



(19)

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

EP 1 225 182 A2

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:  
**24.07.2002 Bulletin 2002/30**

(51) Int Cl.7: **C07K 14/435, C12N 15/63,**  
**C07K 16/18, G01N 33/50,**  
**C12Q 1/68**

(21) Application number: **02250306.4**

(22) Date of filing: **16.01.2002**

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU**  
**MC NL PT SE TR**  
Designated Extension States:  
**AL LT LV MK RO SI**

(30) Priority: **17.01.2001 US 262216**  
**08.05.2001 US 289358**

(71) Applicant: **Millennium Pharmaceuticals, Inc.**  
**Cambridge, MA 02139 (US)**

(72) Inventor: **Curtis, Rory A J**  
**Southborough, MA 01772 Massachusetts (US)**

(74) Representative: **Hall, Robert Leonard et al**  
**Harrison Goddard Foote,**  
**Fountain Precinct,**  
**Leopold Street**  
**Sheffield S1 2QD (GB)**

### (54) Human phospholipid transporter

(57) The invention provides isolated nucleic acids molecules, designated 67073 nucleic acid molecules, which encode novel phospholipid transporter family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 67073 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 67073 gene has been introduced or disrupted. The invention still further provides isolated 67073 proteins, fusion proteins, antigenic peptides and anti-67073 antibodies. Diagnostic and therapeutic methods utilizing compositions of the invention are also provided.



FIGURE 1

**Description**

## Cross-References to Related Applications

5 [0001] This application claims the benefit of U.S. Provisional Application Number 60/262,216, filed January 17, 2001, and U.S. Provisional Application Number 60/289,358, filed May 8, 2001, the contents of which are incorporated herein by this reference.

**Background of the Invention**

10 [0002] Enzymes that bind to and hydrolyze ATP play a pivotal role in translating chemically stored energy into biological activity. ATPases can function in a variety of cellular processes including, selective ion transport events, actin-based motility, membrane traffic and numerous biosynthetic pathways. Multiple ATPase families exist, including ion transporters (often called pumps), DEAD box-helicases, ABC transporters, and AAA (ATPases Associated to a variety of cellular Activities).

15 [0003] Most eukaryotic E1-E2 ATPases reside in the plasma membrane to perform their transport functions. These ATPases typically have ten transmembrane segments with both termini in the cytoplasm (Smith et al., (1993) *J. Biol. Chem.* 268:22469-22479). All members of the E1-E2 ATPase family form an aspartyl-phosphate intermediate in the course of hydrolyzing ATP.

20 [0004] The E1-E2 (or P-type) ATPases are ion transport enzymes and have been identified in prokaryotes and eukaryotes. Most maintain physiological gradients of common cations through a process of active transport, as reviewed in Lehninger, A.H. (1982) *Principles of Biochemistry*, chapter 14. Members of this superfamily include calcium transporting ATPases, sodium/potassium ATPases, and proton ATPases. The calcium pump in muscle cells sequesters calcium in the sarcoplasmic reticulum to effect muscle relaxation. The Na<sup>+</sup>K<sup>+</sup> ATPase of cell membranes maintains a gradient of high sodium, low potassium outside the cell and low sodium, high potassium inside the cell to maintain osmotic balance in the face of high intracellular protein concentration. The proton pump in the parietal cells of the stomach lining maintains stomach pH at 1.0 by the transport of H<sup>+</sup> into the stomach. New members of the E1-E2 ATPase family include aminophospholipid translocases (or phospholipid transport ATPase, Tang et al., (1996) *Science* 272: 1495-1497; Bull, L. N. et al. (1998) *Nat. Genet.* 18:219-224; Mauro, I. et al. (1999) *Biochem. Biophys. Res. Commun.* 257:333-339). These ATPases maintain the lipid composition of cell membranes. With this function, these ATPases participate in membrane recognition events, such as in blood coagulation, apoptosis of lymphocytes, synaptic vesicle fusion and bile acid secretion.

## Summary of the Invention

35 [0005] The present invention is based, in part, on the discovery of a novel phospholipid transporter family member, referred to herein as "67073". The nucleotide sequences of cDNAs encoding 67073 are shown in SEQ ID NOs:1 and 4, and the amino acid sequences of 67073 polypeptides are shown in SEQ ID NOs:2 and 5. In addition, the nucleotide sequences of the coding regions are depicted in SEQ ID NOs:3 and 6.

40 [0006] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 67073 protein or polypeptide, e.g., a biologically active portion of the 67073 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5. In other embodiments, the invention provides isolated 67073 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under a stringent hybridization condition as described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, wherein the nucleic acid encodes a full length 67073 protein or an active fragment thereof.

50 [0007] In a related aspect, the invention further provides nucleic acid constructs which include a 67073 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included are vectors and host cells containing the 67073 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing polypeptides.

55 [0008] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 67073-encoding nucleic acids.

[0009] In still another related aspect, isolated nucleic acid molecules that are antisense to a 67073 encoding nucleic acid molecule are provided.

[0010] In another aspect, the invention features 67073 polypeptides, and biologically active or antigenic fragments

thereof that are useful, *e.g.*, as reagents or targets in assays applicable to treatment and diagnosis of phospholipid transporter-associated or other 67073-associated disorders. In another embodiment, the invention provides 67073 polypeptides having a 67073 activity. Preferred polypeptides are 67073 proteins including at least one E1-E2 ATPase domain and at least one, two, preferably three phospholipid-transporting domains, and, preferably, having a 67073 activity, *e.g.*, a 67073 activity as described herein.

**[0011]** In other embodiments, the invention provides 67073 polypeptides, *e.g.*, a 67073 polypeptide having the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under a stringent hybridization condition as described herein to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, wherein the nucleic acid encodes a full length 67073 protein or an active fragment thereof.

**[0012]** In a related aspect, the invention further provides nucleic acid constructs which include a 67073 nucleic acid molecule described herein.

**[0013]** In a related aspect, the invention provides 67073 polypeptides or fragments operatively linked to non-67073 polypeptides to form fusion proteins.

**[0014]** In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically or selectively bind 67073 polypeptides.

**[0015]** In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 67073 polypeptides or nucleic acids.

**[0016]** In still another aspect, the invention provides a process for modulating 67073 polypeptide or nucleic acid expression or activity, *e.g.*, using the compounds identified in the screens described herein. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 67073 polypeptides or nucleic acids, such as conditions or disorders involving aberrant or deficient phospholipid transporter function or expression. Examples of such disorders include, but are not limited to, cellular membrane-related disorders *e.g.*, blood coagulation disorders; apoptotic disorders, *e.g.*, immune (*e.g.* inflammatory) disorders; hormonal disorders, disorders involving aberrant vesicle function and/or formation, such as neurological disorders (*e.g.* brain disorders, as described below); disorders of inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; and cellular proliferation, growth, or differentiation disorders.

**[0017]** The invention also provides assays for determining the activity of or the presence or absence of 67073 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

**[0018]** In a further aspect, the invention provides assays for determining the presence or absence of a genetic alteration in a 67073 polypeptide or nucleic acid molecule, including for disease diagnosis.

**[0019]** In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, *e.g.*, a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 67073 molecule. In one embodiment, the capture probe is a nucleic acid, *e.g.*, a probe complementary to a 67073 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, *e.g.*, an antibody specific for 67073 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

**[0020]** Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawing

**[0021]** *Figure 1* depicts a hydropathy plot of human 67073. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 67073 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, *e.g.*, a sequence above the dashed line, *e.g.*, the sequence from about amino acid 305 to 326, from about 351 to 370, of SEQ ID NOs:2 and 5, and from about 1017 to 1037 of SEQ ID NO:2; all or part of a hydrophilic sequence, *e.g.*, a sequence below the dashed line, *e.g.*, the sequence from about amino acid 49 to 60, from about 327 to 334, and from about 621 to 629 of SEQ ID NOs:2 and 5; a sequence which includes a Cys, or a glycosylation site.

#### Detailed Description of the Invention

**[0022]** The human 67073 sequence (SEQ ID NO:1), which is approximately 3878 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 3531 nucleotides, including the ter-

mination codon (nucleotides indicated as coding of SEQ ID NO:1; SEQ ID NO:3). The coding sequence encodes a 1176 amino acid protein (SEQ ID NO:2). The partial human 67073 sequence (SEQ ID NO:4), which is approximately 2997 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2652 nucleotides, including the termination codon (nucleotides indicated as coding of SEQ ID NO:4, SEQ ID NO:6). The coding sequence of partial human 67073 encodes an 883 amino acid protein (SEQ ID NO:5).

5 [0023] Human 67073 contains the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997) *Protein* 28:405-420 and

<http://www.psc.edu/general/software/packages/pfam/pfam.html>:

10 [0024] an E1-E2 ATPase domain (PFAM Accession Number PF00122, SEQ ID NO:7) located at about amino acid residues 139 to 182 of SEQ ID NO:2 and SEQ ID NO:5;

[0025] three phospholipid-transporting domains (ProDomain PD004982, SEQ ID NO:8; PD030421, SEQ ID NO:9; and PD038238, SEQ ID NO:10) located at about amino acid residues 199 to 394, 592 to 726 of SEQ ID NO:2 and SEQ ID NO:5, and 1091 to 1176, respectively, of SEQ ID NO:2;

15 [0026] ten transmembrane domains (predicted by MEMSAT, Jones et al. (1994) *Biochemistry* 33:3038-3049) located at about amino acids 73 to 94, 106 to 122, 305 to 326, 351 to 370 of SEQ ID NOs:2 and 5, and 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, and 1052 to 1072 of SEQ ID NO:2;

[0027] one E1-E2 ATPases phosphorylation site (Prosite PS00154, SEQ ID NO:11) located at about amino acids 416 to 422 of SEQ ID NO:2 and SEQ ID NO:5;

20 [0028] one coiled coil structure (PSORT, <http://psort.nibb.ac.jp/>) located at about amino acids 622 to 661 of SEQ ID NO:2 and SEQ ID NO:5;

[0029] three N-glycosylation sites (Prosite PS00001) located at about amino acids 286 to 289, 327 to 330, and 836 to 839 of SEQ ID NO:2 and SEQ ID NO:5;

[0030] three tyrosine kinase phosphorylation sites (Prosite PS00007) located at about amino acids 328 to 336, 502 to 509, and 589 to 596 of SEQ ID NO:2 and SEQ ID NO:5;

25 [0031] thirteen protein kinase C phosphorylation sites (Prosite PS00005) located at about amino acids 63 to 65, 103 to 105, 287 to 289, 415 to 417, 531 to 533, 559 to 561, 572 to 574, 706 to 708, 756 to 758, 775 to 777, of SEQ ID NOs: 2 and 5, and 1083 to 1085, 1168 to 1170, and 1173 to 1175 of SEQ ID NO:2;

30 [0032] sixteen casein kinase II phosphorylation sites (Prosite PS00006) located at about amino acids 34 to 37, 180 to 183, 277 to 280, 295 to 298, 329 to 332, 342 to 345, 449 to 452, 512 to 515, 623 to 626, 665 to 668, 761 to 761, 827 to 830, 846 to 849, of SEQ ID NOs:2 and 5, and 1087 to 1090, 1150 to 1153, and 1158 to 1161 of SEQ ID NO:2;

[0033] four cAMP/cGMP-dependent protein kinase phosphorylation sites (Prosite PS00004) located at about amino acids 139 to 142, 429 to 432, 562 to 565, of SEQ ID NOs:2 and 5, and 1170 to 1173 of SEQ ID NO:2;

35 [0034] two amidation sites (Prosite PS00009) located at about amino acids 1111 to 1114 and 1125 to 1128 of SEQ ID NO:2; and

[0035] fifteen N-myristoylation sites (Prosite PS00008) located at about amino acids 148 to 153, 206 to 211, 224 to 229, 419 to 424, 435 to 440, 544 to 549, 674 to 679, 739 to 744, 815 to 820, 823 to 828, 831 to 836, of SEQ ID NOs: 2 and 5, and 1005 to 1010, 1021 to 1026, 1031 to 1036, and 1142 to 1147 of SEQ ID NO:2.

40 [0036] The 67073 protein contains a significant number of structural characteristics in common with members of the phospholipid transporter family, such as an E1-E2 ATPase domain and phospholipid-transporting domains. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologs of non-human origin, e.g., rat or mouse proteins. Members of a family also can have common functional characteristics.

45 [0037] As used herein, the term "phospholipid transporter" includes a protein or polypeptide which is capable of transporting a molecule, e.g., a phospholipid (e.g., an aminophospholipid such as phosphatidylserine and phosphatidylethanolamine, a choline phospholipid such as phosphatidylcholine and sphingomyelin, and a bile acid) across a membrane, e.g., a cell membrane (e.g., a nerve cell membrane) and, thus, play a role in or function in a variety of cellular processes, e.g., phospholipid transport, absorption, secretion, gene expression, intra- or inter-cellular signaling, and/or cellular proliferation, growth, and/or differentiation. These proteins are new members of the E1-E2 ATPase family which also includes calcium transporting ATPases, sodium/potassium ATPases, and proton ATPases (Tang et al., (1996) *Science* 272: 1495-1497; Bull, L. N. et al. (1998) *Nat. Genet.* 18:219-224; Mauro, I. et al. (1999) *Biochem. Biophys. Res. Commun.* 257:333-339). Members of the phospholipid transporter family of proteins are integral membrane proteins with at least one E1-E2 ATPase domain, at least one, two, preferably three phospholipid-transporting domains, at least one, two, three, four, five, six, seven, eight, nine, preferably ten transmembrane segments and an aspartate residue, which is predicted to be phosphorylated in the course of hydrolyzing ATP. A GAP alignment of the

human 67073 protein with a mouse phospholipid transporter family member, AT1B (P98200 in SwissProt), shows about 94.6 % sequence identity over the 1148 amino acid region of overlap between the two sequences (as calculated by the Blosum 62.ijj matrix).

**[0038]** A 67073 polypeptide can include an "E1-E2 ATPase domain" or regions homologous with an "E1-E2 ATPase domain". A 67073 polypeptide can further include at least one, two, preferably three "phospholipid-transporting domains" or regions homologous with a "phospholipid-transporting domain," and at least one, two, three, four, five, six, seven, eight, nine, preferably ten transmembrane segments.

**[0039]** As used herein, the term "E1-E2 ATPase domain" includes an amino acid sequence of about 20 to 80 amino acid residues in length and having a bit score for the alignment of the sequence to the E1-E2 ATPase domain (HMM) of at least 5. Preferably, an E1-E2 ATPase domain includes at least about 25 to 65 amino acids, more preferably about 35 to 55 amino acid residues, or about 40 to 50 amino acids and has a bit score for the alignment of the sequence to the E1-E2 ATPase domain (HMM) of at least 10, 12, 15 or greater. Preferably an E1-E2 ATPase domain mediates ATP hydrolysis by providing a phosphate acceptor site in the signature sequence, D-K-T-G-T-[LIVM]-[Tl] (SEQ ID NO:11, Prosite entry PS00154). This signature sequence is typically found in the active site of the ATPase catalytic domain and contains a conserved aspartate residue, which is phosphorylated as an intermediate during the catalytic activity of ATP hydrolysis by the enzyme. This signature sequence which is specific for ATPases was identified at about residues 416 to 422 of SEQ ID NO:2 and SEQ ID NO:5. In the above conserved signature sequence, and other motifs or signature sequences described herein, the standard IUPAC one-letter code for the amino acids is used. Each element in the pattern is separated by a dash (-); square brackets ([ ]) indicate the particular residues that are accepted at that position; x indicates that any residue is accepted at that position; and numbers in parentheses (( )) indicate the number of residues represented by the accompanying amino acid. The E1-E2 ATPase domain (HMM) has been assigned the PFAM Accession Number PF00122 (<http://genome.wustl.edu/Pfam/.html>). An alignment of the E1-E2 ATPase domain (amino acids 139 to 182 of SEQ ID NO:2 and SEQ ID NO:5) of human 67073 with the Pfam consensus amino acid sequence (SEQ ID NO:7) derived from a hidden Markov model yielded a bit score of 15.1.

**[0040]** In a preferred embodiment, a 67073 polypeptide or protein has an "E1-E2 ATPase domain" or a region which includes at least about 25 to 65 more preferably about 35 to 55 or 40 to 50 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "E1-E2 ATPase domain," e.g., the E1-E2 ATPase domain of human 67073 (e.g., residues 139 to 182 of SEQ ID NO:2 and SEQ ID NO:5).

**[0041]** To identify the presence of a "E1-E2 ATPase" domain in a 67073 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the Pfam database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters ([http://www.sanger.ac.uk/Software/Pfam/HMM\\_search](http://www.sanger.ac.uk/Software/Pfam/HMM_search)). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of an "E1-E2 ATPase domain" domain in the amino acid sequence of human 67073 at about residues 139 to 182 of SEQ ID NO:2 and SEQ ID NO:5.

**[0042]** The 67073 protein can have at least one, two, preferably three phospholipid-transporting domains. As used herein, the "phospholipid-transporting domains" are amino acid regions which are about 175 to 225 amino acids in length and are homologous, e.g. at least about 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54% or 55% identical to ProDomain family phospholipid-transporting domains (see ProDomain Release 2001.1; <http://www.toulouse.inra.fr/prodom.html>). Preferably, the phospholipid-transporting domains are involved in the regulation of the 67073 protein by interaction with intracellular second messenger-activated signal transduction enzymes. In one embodiment, the 67073 polypeptides of the invention can have three phospholipid-transporting domains. The first phospholipid-transporting domain is homologous to ProDom family PD004982 ("ATPase Hydrolase Transmembrane ATP-Binding Phosphorylation Magnesium 3.6.1-Multigene Potential Phospholipid-Transporting," SEQ ID NO:8). An alignment of the first phospholipid-transporting domain (amino acids 199 to 394 of SEQ ID NO:2 and SEQ ID NO:5) of human 67073 with the PD004982 consensus amino acid sequence (SEQ ID NO:8) derived from BLAST search resulted in 48 % sequence identity. The second phospholipid-transporting domain is homologous to ProDom family PD030421 ("ATPase Hydrolase Transmembrane ATP-Binding Phosphorylation Magnesium 3.6.1-Multigene Phospholipid-Transporting Potential," SEQ ID NO:9). An alignment of the second phospholipid-transporting domain (amino acids 592 to 726 of SEQ ID NO:2 and SEQ ID NO:5) of human 67073 with the PD030421 consensus amino acid sequence (SEQ ID NO:9) derived from BLAST search resulted in 63 % sequence identity. The third phospholipid-transporting domain is homologous to ProDom family PD038238 ("ATPase Family Magnesium 3.6.1-Multigene Phosphorylation Phospholipid-Transporting Hydrolase ATP-Binding Transmembrane," SEQ ID NO:10). An alignment of the

third phospholipid-transporting domain (amino acids 1091 to 1176 of SEQ ID NO:2) of human 67073 with the PD038238 consensus amino acid sequence (SEQ ID NO:10) derived from BLAST search resulted in 74% sequence identity.

**[0043]** To identify the presence of a "phospholipid-transporting" domain in a 67073 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of domains, *e.g.*, the ProDom database (Corpet *et al.* (1999), *Nucl. Acids Res.* 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-BLAST searches (Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Gouzy *et al.* (1999) *Computers and Chemistry* 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain. A BLAST search was performed against the HMM database resulting in the identification of "phospholipid-transporting" domains in the amino acid sequence of human 67073 at about residues 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, and 1091 to 1176, respectively, of SEQ ID NO:2.

**[0044]** A 67073 polypeptide can include at least one, two, three, four, five, six, seven, eight, nine, preferably ten "transmembrane domains" or regions homologous with a "transmembrane domain". As used herein, the term "transmembrane domain" includes an amino acid sequence of about 10 to 40 amino acid residues in length and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, *e.g.*, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains typically have alpha-helical structures and are described in, for example, Zagotta *et al.*, (1996) *Annual Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference. The transmembrane domains of human 67073 are located at about residues 73 to 94, 106 to 122, 305 to 326, 351 to 370, of SEQ ID NOs:2 and 5, and 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, and 1052 to 1072 of SEQ ID NO:2.

**[0045]** In a preferred embodiment, a 67073 polypeptide or protein has at least one, two, three, four, five, six, seven, eight, nine, preferably ten "transmembrane domains" or regions which include at least about 12 to 35, more preferably about 14 to 30 or 15 to 25 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "transmembrane domain," *e.g.*, the transmembrane domains of human 67073 (*e.g.*, residues 73 to 94, 106 to 122, 305 to 326, 351 to 370, of SEQ ID NOs:2 and 5, and 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, and 1052 to 1072 of SEQ ID NO:2). The transmembrane domains of human 67073 are visualized in the hydropathy plot (Figure 1) as regions of about 15 to 25 amino acids where the hydropathy trace is mostly above the horizontal line.

**[0046]** To identify the presence of a "transmembrane" domain in a 67073 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be analyzed by a transmembrane prediction method that predicts the secondary structure and topology of integral membrane proteins based on the recognition of topological models (MEMSAT, Jones *et al.*, (1994) *Biochemistry* 33: 3038-3049).

**[0047]** A 67073 polypeptide can include at least one, two, three, four, five, six, seven, eight, nine, ten, preferably eleven "non-transmembrane regions." As used herein, the term "non-transmembrane region" includes an amino acid sequence not identified as a transmembrane domain. The non-transmembrane regions in 67073 are located at about amino acids 1 to 72, 95 to 105, 123 to 304, 327 to 350, and SEQ ID NOs:2 and 5, and 371 to 874, 892 to 898, 920 to 947, 967 to 984, 1003 to 1016, 1038 to 1051, and 1073 to 1176 of SEQ ID NO:2.

**[0048]** The non-transmembrane regions of 67073 include at least one, two, three, four, five, preferably six cytoplasmic regions. When located at the N-terminus, the cytoplasmic region is referred to herein as the "N-terminal cytoplasmic domain." As used herein, an "N-terminal cytoplasmic domain" includes an amino acid sequence having about 1 to 125, preferably about 1 to 100, more preferably about 1 to 85, or even more preferably about 1 to 75 amino acid residues in length and is located inside of a cell or within the cytoplasm of a cell. The C-terminal amino acid residue of an "N-terminal cytoplasmic domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a 67073 protein. For example, an N-terminal cytoplasmic domain is located at about amino acid residues 1 to 72 of SEQ ID NO:2 and SEQ ID NO:5.

**[0049]** In a preferred embodiment, a 67073 polypeptide or protein has an N-terminal cytoplasmic domain or a region which includes about 1 to 125, preferably about 1 to 85, and more preferably about 1 to 75 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with an "N-terminal cytoplasmic domain," *e.g.*, the N-terminal cytoplasmic domain of human 67073 (*e.g.*, residues 1 to 72 of SEQ ID NO:2 and SEQ ID NO:5).

**[0050]** In another embodiment, the 67073 cytoplasmic regions include at least one, two, three, preferably four cytoplasmic loops. As used herein, the term "loop" includes an amino acid sequence which is not included within a phospholipid membrane, having a length of at least about 4, preferably about 5 to 550, more preferably about 6 to 510 amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a 67073 molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a

transmembrane domain in a 67073 molecule. As used herein, a "cytoplasmic loop" includes a loop located inside of a cell or within the cytoplasm of a cell. For example, a "cytoplasmic loop" can be found at about amino acid residues 123 to 304, of SEQ ID NOs:2 and 5, and 371 to 874, 920 to 947, and 1003 to 1016 of SEQ ID NO:2.

**[0051]** In a preferred embodiment, a 67073 polypeptide or protein has a cytoplasmic loop or a region which includes at least about 6, preferably about 10 to 550, and more preferably about 12 to 510 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a cytoplasmic loop," *e.g.*, a cytoplasmic loop of human 67073 (*e.g.*, residues 123 to 304, of SEQ ID NOs:2 and 5, and 371 to 874, 920 to 947, and 1003 to 1016 of SEQ ID NO:2).

**[0052]** In another embodiment, the 67073 non-transmembrane regions include at least one, two, three, four, preferably five non-cytoplasmic loops. As used herein, a "non-cytoplasmic loop" includes an amino acid sequence located outside of a cell or within an intracellular organelle. Non-cytoplasmic loops include extracellular domains (*i.e.*, outside of the cell) and intracellular domains (*i.e.*, within the cell). When referring to membrane-bound proteins found in intracellular organelles (*e.g.*, mitochondria, endoplasmic reticulum, peroxisomes microsomes, vesicles, endosomes, and lysosomes), non-cytoplasmic loops include those domains of the protein that reside in the lumen of the organelle or the matrix or the intermembrane space. Members of the phospholipid transporter family typically reside on the plasma membrane, so the non-cytoplasmic loops are "extracellular." For example, a "non-cytoplasmic or extracellular loop" can be found at about amino acid residues 95 to 105, 327 to 350, of SEQ ID NOs:2 and 5, and 892 to 898, 967 to 984, and 1038 to 1051 of SEQ ID NO:2.

**[0053]** In a preferred embodiment, a 67073 polypeptide or protein has at least one non-cytoplasmic loop or a region which includes at least about 4, preferably about 5 to 50, more preferably about 6 to 25 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "non-cytoplasmic or extracellular loop," *e.g.*, at least one non-cytoplasmic or extracellular loop of human 67073 (*e.g.*, residues 95 to 105, 327 to 350, of SEQ ID NOs:2 and 5, and 892 to 898, 967 to 984, and 1038 to 1051 of SEQ ID NO:2).

**[0054]** In another embodiment, a cytoplasmic region of a 67073 protein can include the C-terminus and can be a "C-terminal cytoplasmic domain," also referred to herein as a "C-terminal cytoplasmic tail." As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 50, preferably about 75 to 125, more preferably about 100 to 110 amino acid residues and is located inside of a cell or within the cytoplasm of a cell. The N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a 67073 protein. For example, a C-terminal cytoplasmic domain is located at about amino acid residues 1073 to 1176 of SEQ ID NO:2.

**[0055]** In a preferred embodiment, a 67073 polypeptide or protein has a C-terminal cytoplasmic domain or a region which includes at least about 50, preferably about 75 to 125, more preferably about 100 to 110 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a C-terminal cytoplasmic domain," *e.g.*, the C-terminal cytoplasmic domain of human 67073 (*e.g.*, residues 1073 to 1176 of SEQ ID NO:2).

**[0056]** A 67073 phospholipid transporter also can include a coiled coil structure. Coiled coil structures are supercoiled helical domains responsible for the oligomerization of proteins. There is a characteristic heptad repeat (h-x-x-h-x-x-x) n in the coiled coil structures, where h represents hydrophobic residues (Beck and Brodsky (1998) *J. Struct. Biol.* 122: 17-29). Coiled coil structures are found in a wide variety of proteins, including cytoskeletal, nuclear, muscle, cell surface, extracellular, plasma, bacterial, and viral proteins and can be found in the 67073 polypeptide at about amino acids 622 to 661 of SEQ ID NO:2 and SEQ ID NO:5.

**[0057]** A 67073 family member can include at least one E1-E2 ATPase domain; at least one, two, preferably three phospholipid-transporting domains, at least one, two, three, four, five, six, seven, eight, nine, preferably ten transmembrane domains. A 67073 family member also can include at least one E1-E2 ATPase phosphorylation site (Prosite PS00154) and a coiled coil structure (PSORT, <http://psort.nibb.ac.jp/>). Furthermore, a 67073 family member can include at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, preferably thirteen protein kinase C phosphorylation sites (Prosite PS00005); at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, preferably sixteen casein kinase II phosphorylation sites (Prosite PS00006); at least one, two, preferably three N-glycosylation sites (Prosite PS00001); at least one, two, three, preferably four cAMP/cGMP protein kinase phosphorylation sites (Prosite PS00004); at least one, two, preferably three tyrosine kinase phosphorylation sites (Prosite PS00007); at least one, preferably two amidation sites (Prosite PS00009); and at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, and preferably fifteen N-myristylation sites (Prosite PS00008).

**[0058]** As the 67073 polypeptides of the invention can modulate 67073-mediated activities, they can be useful for developing novel diagnostic and therapeutic agents for phospholipid transporter-associated or other 67073-associated disorders, as described below.

**[0059]** As used herein, a "phospholipid transporter-associated activity" includes an activity which involves participation in regulating a variety of cellular processes, *e.g.*, hydrolysis of ATP, phospholipid transport (*e.g.*, aminophospholipid transport), absorption, secretion, gene expression, intra- or inter-cellular signaling, and/or cellular proliferation, growth, apoptosis, and/or differentiation. These transporters maintain the lipid composition of cell membranes. With this func-

tion, these transporters participate in membrane recognition events, such as in blood coagulation, apoptosis of lymphocytes, synaptic vesicle fusion and bile acid secretion.

**[0060]** As used herein, a "67073 activity", "biological activity of 67073" or "functional activity of 67073", refers to an activity exerted by a 67073 protein, polypeptide or nucleic acid molecule on *e.g.*, a 67073-responsive cell or on a 67073 substrate, *e.g.*, a protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 67073 activity is a direct activity, such as an association with a 67073 target molecule. A "target molecule" or "binding partner" is a molecule with which a 67073 protein binds or interacts in nature. In an exemplary embodiment, 67073 is a transporter, *e.g.*, an E1-E2 ATPase phospholipid transporter family member, and thus binds to or interacts in nature with molecule, *e.g.*, a phospholipid (*e.g.*, an aminophospholipid such as phosphatidylserine and phosphatidylethanolamine, a choline phospholipid such as phosphatidylcholine and sphingomyelin, and a bile acid).

**[0061]** A 67073 activity can also be an indirect activity, *e.g.*, a cellular signaling activity mediated by interaction of the 67073 protein with a 67073 receptor. Based on the above-described sequence structures and similarities to molecules of known function, the 67073 molecules of the present invention can have similar biological activities as phospholipid transporter family members. For example, the 67073 proteins of the present invention can have one or more of the following activities: (1) the ability to anchor into a membrane, *e.g.*, a cell membrane (*e.g.*, a nerve cell membrane); (2) the ability to be transiently phosphorylated on an acidic (*e.g.*, aspartate) residue; (3) the ability to modulate the phosphorylation state of a 67073 target molecule (*e.g.*, an ATP molecule) by, *e.g.*, inducing ATP hydrolysis; (4) the ability to maintain a gradient, *e.g.*, a molecular gradient (*e.g.*, a phospholipid gradient), across a membrane; (5) the ability to transport phospholipids (*e.g.*, aminophospholipids such as phosphatidylserine and phosphatidylethanolamine, choline phospholipids such as phosphatidylcholine and sphingomyelin, and bile acids) across a membrane; (6) the ability to modulate the location of a substrate or target molecule (*e.g.*, modulation of phospholipid location within a cell and/or location with respect to a cellular membrane); (7) the ability to modulate apoptosis or cell death; (8) the ability to modulate immune cell activity; (9) the ability to modulate blood coagulation; (10) the ability to modulate energy metabolism; (11) the ability to modulate organ function; (12) the ability to modulate neuronal activity, *e.g.*, neuron stimulation; (13) the ability to modulate the activity of one or more proteins involved in cellular growth, proliferation, or differentiation, *e.g.*, cardiac, epithelial, or neuronal cell growth, proliferation or differentiation; (14) the ability to modulate cancer or tumor progression; and (15) the ability to modulate tissue development (*e.g.* embryogenesis). Thus, the 67073 molecules of the present invention can participate in one or more of the following processes: (1) the regulation of membrane phospholipid composition; (2) the modulation of blood coagulation; (3) the modulation of cellular differentiation, proliferation, growth, absorption or secretion; (4) modulation of cell death; and (5) the regulation of metabolic pathways.

**[0062]** As used herein, a "phospholipid transporter-associated disorder", includes a disorder, disease, or condition characterized by a deregulation, *e.g.*, an upregulation, a downregulation, or an abnormal function of a phospholipid transporter involved in cellular membrane structure. In a preferred embodiment, a phospholipid transporter-associated disorder includes the inhibition or over-stimulation of the activity of phospholipid transporter involved in, *e.g.*, phospholipid transport associated with cellular membrane formation and/or maintenance. Such fluctuations in membrane composition can lead to perturbed membrane structure, which can in turn lead membrane structure-related disorders. Cellular membrane structure deregulation may be due to, or result in, a deregulation or misregulation of cellular proliferation, cell cycle progression, cellular differentiation (*e.g.*, cellular growth related disorders), cell metabolism, cellular hypertrophy, inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; systemic responses in an organism, such as nervous system responses, hormonal responses (*e.g.*, insulin response), or immune responses; and protection of cells from toxic compounds (*e.g.*, carcinogens, toxins, or mutagens).

**[0063]** Thus, the 67073 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more phospholipid transporter, *e.g.*, cellular membrane-related disorders. Examples of such disorders, *e.g.*, phospholipid transporter-associated or other 67073-associated disorders, include but are not limited to, blood coagulation disorders; apoptotic disorders; immune (*e.g.* inflammatory) disorders; hormonal disorders, disorders involving aberrant vesicle function and/or formation, such as neurological disorders (*e.g.* brain disorders, as described below); disorders of inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; cellular proliferation, growth, or differentiation disorders and viral diseases. Examples of cellular growth related disorders include cardiovascular disorders, such as heart failure, dilated cardiomyopathy, idiopathic cardiomyopathy, or angina; proliferative disorders or differentiative disorders, such as cancer, *e.g.*, melanoma, prostate cancer, cervical cancer, breast cancer, colon cancer, sarcoma, lymphoma or leukemia. Examples of cell-cell communication or cell metabolism disorders include neurological disorders (*e.g.* brain disorders, as described below, including mitochondrial encephalomyopathies); leukodystrophies; immune or inflammatory disorders; and metabolic ion balance disorders, such as cystic fibrosis, hypertension or atrial fibrillation. These and other disorders are described in more detail below.

**[0064]** The 67073 molecules can be used to treat blood coagulation disorders, in part because phospholipid transporters are involved in membrane recognition events such as blood coagulation. Blood coagulation disorders include, but are not limited to, disorders resulting from aberrant or deficient platelet function, such as thrombocytopenia, Ber-

nard-Soulier syndrome or thrombasthenia, disorders involving aberrant or deficient clotting factors, such as hemophilia or von Willebrand's disease, or disorders resulting in disseminated intravascular coagulation, such as obstetric complications, infections, neoplastic diseases or traumatic tissue injuries.

**[0065]** The 67073 molecules can be used to treat apoptotic disorders, in part because phospholipid transporters are involved in apoptosis. Disorders involving aberrant or deficient apoptosis include, but are not limited to, autoimmune disorders such as systemic lupus erythematosus and immune-mediated glomerulonephritis; neoplastic disorders such as follicular lymphoma and hormone dependent tumors of the breast, prostate gland and ovary; neurodegenerative disorders, such as Alzheimer's disease, Huntington's disease, retinitis pigmentosa, amyotrophic lateral sclerosis, spinal muscular atrophy and Parkinson's disease; viral infections, such as those caused by herpesviruses, poxviruses and adenoviruses; blood disorders due to aberrant apoptotic activity in the bone marrow, such as anemia associated with chronic disease, *i.e.*, aplastic anemia, chronic neutropenia and myelodysplasia; and tissue damage associated with myocardial infarction and stroke.

**[0066]** The 67073 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, *e.g.*, inflammatory, (*e.g.* respiratory inflammatory) disorders. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease, *e.g.* Crohn's disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

**[0067]** Phospholipid transporter-associated disorders can include hormonal disorders, such as conditions or diseases in which the production, secretion and/or regulation of hormones in an organism is aberrant or deficient. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

**[0068]** The 67073 molecules can be used to treat neurological disorders, in part because phospholipid transporters are involved in membrane specialization and/or recognition events such as synaptic vesicle formation and vesicle fusion. Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme's disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer's disease and Pick's disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson's disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington's disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich's ataxia, and ataxiatelangiect-

sia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe's disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan's disease, mitochondrial encephalomyopathies, including Leigh's disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B<sub>1</sub>) deficiency and vitamin B<sub>12</sub> deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephalopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendrogloma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

**[0069]** The 67073 molecules can be used to treat cardiovascular disorders, in part because phospholipid transporters are involved in membrane recognition events in cardiovascular tissues. As used herein, disorders involving the heart, or "cardiovascular disease" or a "cardiovascular disorder" includes a disease or disorder which affects the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, e.g., by a thrombus. A cardiovascular disorder includes, but is not limited to disorders such as arteriosclerosis, atherosclerosis, cardiac hypertrophy, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, valvular disease, including but not limited to, valvular degeneration caused by calcification, rheumatic heart disease, endocarditis, or complications of artificial valves; atrial fibrillation, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, pericardial disease, including but not limited to, pericardial effusion and pericarditis; cardiomyopathies, e.g., dilated cardiomyopathy or idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, ischemic disease, arrhythmia, sudden cardiac death, and cardiovascular developmental disorders (e.g., arteriovenous malformations, arteriovenous fistulae, raynaud's syndrome, neurogenic thoracic outlet syndrome, causalgia/reflex sympathetic dystrophy, hemangioma, aneurysm, cavernous angioma, aortic valve stenosis, atrial septal defects, atrioventricular canal, coarctation of the aorta, ebsteins anomaly, hypoplastic left heart syndrome, interruption of the aortic arch, mitral valve prolapse, ductus arteriosus, patent foramen ovale, partial anomalous pulmonary venous return, pulmonary atresia with ventricular septal defect, pulmonary atresia without ventricular septal defect, persistence of the fetal circulation, pulmonary valve stenosis, single ventricle, total anomalous pulmonary venous return, transposition of the great vessels, tricuspid atresia, truncus arteriosus, ventricular septal defects). A cardiovascular disease or disorder also can include an endothelial cell disorder.

**[0070]** As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, e.g., proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (e.g., atherosclerosis), and chronic inflammatory diseases (e.g., rheumatoid arthritis).

**[0071]** Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

**[0072]** As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, e.g., malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, e.g., cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary sys-

tem carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

**[0073]** The 67073 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus (1991) *Crit Rev. in Oncol./Hemotol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

**[0074]** The 67073 molecules can be used to treat viral diseases, in part because phospholipid transporters are involved in the apoptosis associated with viral infections. These viral diseases, include but are not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 67073 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 67073 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

**[0075]** The 67073 molecules can be used to treat testicular disorders, in part because these molecules can be isolated from a testis cDNA library and in part because of the membrane recognition events required for proper function of the testis and the blood-testis barrier. Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumor of sex cord-gonadal stroma including, but not limited to, Leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

**[0076]** The 67073 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 or SEQ ID NO:5 thereof are collectively referred to as "polypeptides or proteins of the invention" or "67073 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "67073 nucleic acids."

**[0077]** As used herein, the term "nucleic acid molecule" includes DNA molecules (*e.g.*, a cDNA or genomic DNA) and RNA molecules (*e.g.*, an mRNA) and analogs of the DNA or RNA generated, *e.g.*, by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

**[0078]** The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

**[0079]** As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in *Current Protocols in Molecular Biology* (1989) John Wiley & Sons, N.Y., 6.3.1-6.3.6, which is incorporated by reference. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6X

sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions); 2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred conditions and the ones that should be used unless otherwise specified.

**[0080]** As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

**[0081]** As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 67073 protein, preferably a mammalian 67073 protein, and can further include non-coding regulatory sequences, and introns.

**[0082]** An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 67073 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-67073 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-67073 chemicals. When the 67073 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

**[0083]** A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 67073 (e.g., the sequence of SEQ ID NO:1, 3, 4 or 6) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the E1-E2 ATPase domain, are predicted to be particularly unamenable to alteration.

**[0084]** A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 67073 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 67073 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 67073 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, SEQ ID NO:3 SEQ ID NO:4, SEQ ID NO:6, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

**[0085]** As used herein, a "biologically active portion" of a 67073 protein includes a fragment of a 67073 protein which participates in an interaction between a 67073 molecule and a non-67073 molecule. Biologically active portions of a 67073 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 67073 protein, e.g., the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:5, which include fewer amino acids than the full length 67073 protein, and exhibit at least one activity of a 67073 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 67073 protein, e.g., hydrolysis of ATP or phospholipid transport. A biologically active portion of a 67073 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 67073 protein can be used as targets for developing agents which modulate a 67073 mediated activity, e.g., hydrolysis of ATP or phospholipid transport.

**[0086]** Calculations of homology or sequence identity (the terms "homology" and "identity" are used interchangeably herein) between sequences are performed as follows:

**[0087]** To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 67073 amino acid sequence of SEQ ID NO:2 having 1176 amino acid residues, at least [30%] 352, preferably at least [40%] 470, more preferably at least [50%] 588, even more preferably at least [60%] 704,

and even more preferably at least [70%] 823, [80%] 940, or [90%] 1058 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

**[0088]** The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

**[0089]** The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers and Miller ((1989) CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

**[0090]** The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 67073 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 67073 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

**[0091]** Particular 67073 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:2 or SEQ ID NO:5 are termed substantially identical.

**[0092]** In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:1, 3, 4 or 6 are termed substantially identical.

**[0093]** "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

**[0094]** "Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

[0095] A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[0096] Various aspects of the invention are described in further detail below. Isolated Nucleic Acid Molecules

[0097] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 67073 polypeptide described herein, *e.g.*, a full length 67073 protein or a fragment thereof, *e.g.*, a biologically active portion of 67073 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, *e.g.*, to identify a nucleic acid molecule encoding a polypeptide of the invention, 67073 mRNA, and fragments suitable for use as primers, *e.g.*, PCR primers for the amplification or mutation of nucleic acid molecules.

[0098] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or SEQ ID NO:4 or a portion of any of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 67073 protein (*i.e.*, "the coding region" of SEQ ID NO: 1 or SEQ ID NO:4, as shown in SEQ ID NO:3 or SEQ ID NO:6, respectively), as well as 5' untranslated sequences (nucleotides 1 to 19 of SEQ ID NO:1 and SEQ ID NO:4) and 3' untranslated sequences (nucleotides 3552 to 3878 of

SEQ ID NO:1 or 2672 to 2997 of SEQ ID NO:4). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (*e.g.*, SEQ ID NO:3) or SEQ ID NO:4 (*e.g.* SEQ ID NO:6) and, *e.g.*, no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to a fragment of the protein from about amino acid 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5.

[0099] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 6, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, 3, 4 or 6, thereby forming a stable duplex.

[0100] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about: 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a portion, preferably of the same length, of any of these nucleotide sequences.

### 30 67073 Nucleic Acid Fragments

[0101] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1, 3, 4, or 6. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 67073 protein, *e.g.*, an immunogenic or biologically active portion of a 67073 protein. A fragment can comprise those nucleotides of SEQ ID NO:1 or SEQ ID NO:4, which encode an E1-E2 ATPase domain of human 67073. The nucleotide sequence determined from the cloning of the 67073 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 67073 family members, or fragments thereof, as well as 67073 homologs, or fragments thereof, from other species.

[0102] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 40 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[0103] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, a 67073 nucleic acid fragment can include a sequence corresponding to an E1-E2 ATPase domain, as described herein.

[0104] 67073 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6.

[0105] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered

differences.

**[0106]** A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes: an E1-E2 ATPase domain from about amino acid 139 to 182 of SEQ ID NOs:2 and 5, a phospholipid-transporting domain located at about amino acid residues 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, or 1091 to 1176 of SEQ ID NO:2; or a transmembrane domain located at about amino acids 73 to 94, 106 to 122, 305 to 326, 351 to 370, of SEQ ID NOs:2 and 5, and 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, and 1052 to 1072 of SEQ ID NO:2.

**[0107]** In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of a 67073 sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: an E1-E2 ATPase domain from about amino acid 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5; a phospholipid-transporting domain located at about amino acid residues 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, or 1091 to 1176 of SEQ ID NO:2; or a transmembrane domain located at about amino acids 73 to 94, 106 to 122, 305 to 326, 351 to 370, of SEQ ID NOs:2 and 5, or 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, and 1052 to 1072 of SEQ ID NO:2. A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

**[0108]** A nucleic acid fragment encoding a "biologically active portion of a 67073 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NOs:1, 3, 4, or 6 which encodes a polypeptide having a 67073 biological activity (*e.g.*, the biological activities of the 67073 proteins are described herein), expressing the encoded portion of the 67073 protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 67073 protein. For example, a nucleic acid fragment encoding a biologically active portion of 67073 includes an E1-E2 ATPase domain, *e.g.*, amino acid residues about 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5. A nucleic acid fragment encoding a biologically active portion of a 67073 polypeptide, can comprise a nucleotide sequence which is greater than 120 or more nucleotides in length.

**[0109]** In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800 or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6. 67073 Nucleic Acid Variants

**[0110]** The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 67073 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2 or SEQ ID NO:5. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

**[0111]** Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. *E.g.*, the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

**[0112]** Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

**[0113]** In a preferred embodiment, the nucleic acid differs from that of SEQ ID NOs:1, 3, 4, or 6, *e.g.*, as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the nucleotides in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

**[0114]** Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in SEQ ID NO:2, SEQ ID NO:5, or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO 2, SEQ ID NO:5, or a fragment of the sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 67073 cDNAs of the invention can further be isolated by mapping to the same chromosome

or locus as the 67073 gene.

**[0115]** Preferred variants include those that are correlated with hydrolysis of ATP or phospholipid transport.

**[0116]** Allelic variants of 67073, *e.g.*, human 67073, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 67073 protein within a population that maintain the ability to bind and hydrolyze ATP or bind and transport phospholipids. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, SEQ ID NO:5, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 67073, *e.g.*, human 67073, protein within a population that do not have the ability to bind and hydrolyze ATP or bind and transport phospholipids. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

**[0117]** Moreover, nucleic acid molecules encoding other 67073 family members and, thus, which have a nucleotide sequence which differs from the 67073 sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6 are intended to be within the scope of the invention.

#### Antisense Nucleic Acid Molecules, Ribozymes and Modified 67073 Nucleic Acid Molecules

**[0118]** In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 67073. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 67073 coding strand, or to only a portion thereof (*e.g.*, the coding region of human 67073 corresponding to SEQ ID NO:3 or SEQ ID NO:6). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 67073 (*e.g.*, the 5' and 3' untranslated regions).

**[0119]** An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 67073 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 67073 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 67073 mRNA, *e.g.*, between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

**[0120]** An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

**[0121]** The antisense nucleic acid molecules of the invention are typically administered to a subject (*e.g.*, by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 67073 protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically or selectively bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

**[0122]** In yet another embodiment, the antisense nucleic acid molecule of the invention is an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

**[0123]** In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 67073-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 67073 cDNA disclosed herein (*i.e.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff

and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 67073-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 67073 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) *Science* 261:1411-1418.

**[0124]** 67073 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 67073 (e.g., the 67073 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 67073 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des.* 6:569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

**[0125]** The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

**[0126]** A 67073 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci.* 93: 14670-675.

**[0127]** PNAs of 67073 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antogene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 67073 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup *et al.* (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

**[0128]** In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

**[0129]** The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 67073 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 67073 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

#### 45 Isolated 67073 Polypeptides

**[0130]** In another aspect, the invention features, an isolated 67073 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-67073 antibodies. 67073 protein can be isolated from cells or tissue sources using standard protein purification techniques. 67073 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

**[0131]** Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present in a native cell.

**[0132]** In a preferred embodiment, a 67073 polypeptide has one or more of the following characteristics:

**[0133]** it has the ability to anchor into a membrane, e.g., a cell membrane (e.g., a nerve cell membrane);

**[0134]** it has the ability to be transiently phosphorylated on an acidic (e.g., aspartate) residue;

[0135] it has the ability to modulate the phosphorylation state of a 67073 target molecule (e.g., an ATP molecule) by, e.g., inducing ATP hydrolysis;

[0136] it has the ability to maintain a gradient, e.g., a molecular gradient (e.g., a phospholipid gradient), across a membrane;

5 [0137] it has the ability to transport phospholipids (e.g., aminophospholipids such as phosphatidylserine and phosphatidylethanolamine, choline phospholipids such as phosphatidylcholine and sphingomyelin, and bile acids) across a membrane;

[0138] it has the ability to modulate the location of a substrate or target molecule (e.g., modulation of phospholipid location within a cell and/or location with respect to a cellular membrane);

10 [0139] it has the ability to modulate apoptosis or cell death;

[0140] it has the ability to modulate immune cell activity;

[0141] it has a molecular weight, e.g., a deduced molecular weight, preferably ignoring any contribution of post translational modifications, amino acid composition or other physical characteristic of a 67073 polypeptide, e.g., a polypeptide of SEQ ID NO:2 or SEQ ID NO:5;

15 [0142] it has an overall sequence similarity of at least 60%, preferably at least 70%, more preferably at least 80, 90, or 95%, with a polypeptide of SEQ ID NO:2 or SEQ ID NO:5; it can be found in spinal cord and testis;

[0143] it has an E1-E2 ATPase domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5;

20 [0144] it has a phospholipid-transporting domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, or 1091 to 1176, respectively, of SEQ ID NO:2; and

[0145] it has a transmembrane domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 73 to 94, 106 to 122, 305 to 326, 351 to 370, of SEQ ID NOs:2 and 5, or 875 to 891, 899 to 919, 948 to 966, 985 to 1002, 1017 to 1037, or 1052 to 1072 of SEQ ID NO:2.

[0146] In a preferred embodiment the 67073 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2 or SEQ ID NO:5. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, 25 preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the E1-E2 ATPase domain at about residues 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5 nor in a phospholipid-transporting domain at about amino acid residues about 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, or 1091 to 1176, respectively, of SEQ ID NO:2. In another embodiment one or more differences are in the E1-E2 ATPase domain at about residues 139 to 182 of SEQ ID NO:2 or SEQ ID NO:5 or in a phospholipid-transporting 30 domain at about amino acid residues about 199 to 394, 592 to 726 of SEQ ID NOs:2 and 5, or 1091 to 1176, respectively, of SEQ ID NO:2.

[0147] Other embodiments include a protein that contains one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity. Such 67073 proteins differ in amino acid sequence from SEQ ID NO:2 or SEQ ID NO:5, yet retain biological activity.

40 [0148] In one embodiment, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more homologous to SEQ ID NO:2 or SEQ ID NO:5.

[0149] A 67073 protein or fragment is provided which varies from the sequence of SEQ ID NO:2 or SEQ ID NO:5 in regions defined by amino acids about 1 to 138, 183 to 198, 395 to 591, or SEQ ID NO:2 by about amino acids 789 to 45 1090 by at least one but by less than 15, 10 or 5 amino acid residues in the protein or fragment but which does not differ from SEQ ID NO:2 or SEQ ID NO:5 in regions defined by amino acids about 139 to 182, 199 to 394, 592 to 726 or SEQ ID NO:2 by about amino acids 1091 to 1176. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

50 [0150] In one embodiment, a biologically active portion of a 67073 protein includes an E1-E2 ATPase domain. In another embodiment, a biologically active portion of a 67073 protein includes a phospholipid-transporting domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 67073 protein.

[0151] In a preferred embodiment, the 67073 protein has an amino acid sequence shown in SEQ ID NO:2 or SEQ 55 ID NO:5. In other embodiments, the 67073 protein is sufficiently or substantially identical to SEQ ID NO:2 or SEQ ID NO:5. In yet another embodiment, the 67073 protein is sufficiently or substantially identical to SEQ ID NO:2 or SEQ ID NO:5 and retains the functional activity of the protein of SEQ ID NO:2 or SEQ ID NO:5, as described in detail in the subsections above.

## 67073 Chimeric or Fusion Proteins

**[0152]** In another aspect, the invention provides 67073 chimeric or fusion proteins. As used herein, a 67073 "chimeric protein" or "fusion protein" includes a 67073 polypeptide linked to a non-67073 polypeptide. A "non-67073 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 67073 protein, *e.g.*, a protein which is different from the 67073 protein and which is derived from the same or a different organism. The 67073 polypeptide of the fusion protein can correspond to all or a portion *e.g.*, a fragment described herein of a 67073 amino acid sequence. In a preferred embodiment, a 67073 fusion protein includes at least one (or two) biologically active portion of a 67073 protein. The non-67073 polypeptide can be fused to the N-terminus or C-terminus of the 67073 polypeptide.

**[0153]** The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-67073 fusion protein in which the 67073 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 67073. Alternatively, the fusion protein can be a 67073 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of 67073 can be increased through use of a heterologous signal sequence.

**[0154]** Fusion proteins can include all or a part of a serum protein, *e.g.*, a portion of an immunoglobulin (*e.g.*, IgG, IgA, or IgE), *e.g.*, an Fc region and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

**[0155]** The 67073 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 67073 fusion proteins can be used to affect the bioavailability of a 67073 substrate. 67073 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 67073 protein; (ii) mis-regulation of the 67073 gene; and (iii) aberrant post-translational modification of a 67073 protein.

**[0156]** Moreover, the 67073-fusion proteins of the invention can be used as immunogens to produce anti-67073 antibodies in a subject, to purify 67073 ligands and in screening assays to identify molecules which inhibit the interaction of 67073 with a 67073 substrate.

**[0157]** Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A 67073-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 67073 protein. Variants of 67073 Proteins

**[0158]** In another aspect, the invention also features a variant of a 67073 polypeptide, *e.g.*, which functions as an agonist (mimetics) or as an antagonist. Variants of the 67073 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation, the insertion or deletion of sequences or the truncation of a 67073 protein. An agonist of the 67073 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 67073 protein. An antagonist of a 67073 protein can inhibit one or more of the activities of the naturally occurring form of the 67073 protein by, for example, competitively modulating a 67073-mediated activity of a 67073 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 67073 protein.

**[0159]** Variants of a 67073 protein can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a 67073 protein for agonist or antagonist activity.

**[0160]** Libraries of fragments *e.g.*, N terminal, C terminal, or internal fragments, of a 67073 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 67073 protein.

**[0161]** Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

**[0162]** Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 67073 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6:327-331).

**[0163]** Cell based assays can be exploited to analyze a variegated 67073 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, a cell line, which ordinarily responds to 67073 in a substrate-dependent manner. The transfected cells are then contacted with 67073 and the effect of the expression of the mutant on signaling by the 67073 substrate can be detected, *e.g.*, by measuring the ability of 67073 to reside within a membrane, hydrolyze ATP or transport a phospholipid. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 67073 substrate, and the individual clones further characterized.

**[0164]** In another aspect, the invention features a method of making a 67073 polypeptide, *e.g.*, a peptide having a non-wild type activity, *e.g.*, an antagonist, agonist, or super agonist of a naturally occurring 67073 polypeptide, *e.g.*, a naturally occurring 67073 polypeptide. The method includes altering the sequence of a 67073 polypeptide, *e.g.*, altering

the sequence, *e.g.*, by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

**[0165]** In another aspect, the invention features a method of making a fragment or analog of a 67073 polypeptide a biological activity of a naturally occurring 67073 polypeptide. The method includes altering the sequence, *e.g.*, by substitution or deletion of one or more residues, of a 67073 polypeptide, *e.g.*, altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

#### Anti-67073 Antibodies

**[0166]** In another aspect, the invention provides an anti-67073 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, *i.e.*, an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, Fab and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as papain or pepsin, respectively.

**[0167]** The antibody can be a polyclonal, monoclonal, recombinant, *e.g.*, a chimeric or humanized, fully human, non-human, *e.g.*, murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

**[0168]** A full-length 67073 protein or, antigenic peptide fragment of 67073 can be used as an immunogen or can be used to identify anti-67073 antibodies made with other immunogens, *e.g.*, cells, membrane preparations, and the like. The antigenic peptide of 67073 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:5 and encompasses an epitope of 67073. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

**[0169]** Fragments of 67073 which include residues about 49 to 60, from about 327 to 334, and from about 621 to 629 of SEQ ID NO:2 or SEQ ID NO:5 can be used to make, *e.g.*, used as immunogens or used to characterize the specificity of an antibody, antibodies against hydrophilic regions of the 67073 protein (see Figure 1). Similarly, fragments of 67073 which include residues about 305 to 326, from about 351 to 370 of SEQ ID NOs:2 and 5, and from about 1017 to 1037 of SEQ ID NO:2 can be used to make an antibody against a hydrophobic region of the 67073 protein; fragments of 67073 which include residues about 327 to 350, or a subset thereof, *e.g.* about residues 327 to 337, or about residues 338 to 350 of SEQ ID NOs:2 and 5, or a fragment which includes residues about 95 to 105 of SEQ ID NOs:2 and 5, or about 1038 to 1051 of SEQ ID NO:2 can be used to make an antibody against an non-cytoplasmic or extracellular region of the 67073 protein; fragments of 67073 which include residues about 1 to 50, about 200 to 300, or about 400 to 550 of SEQ ID NO:2 or SEQ ID NO:5 can be used to make an antibody against an intracellular region of the 67073 protein; a fragment of 67073 which include residues about 139 to 149, about 150 to 164, or about 165 to 182 of SEQ ID NO:2 or SEQ ID NO:5 can be used to make an antibody against the E1-E2 ATPase region of the 67073 protein; a fragment of 67073 which include residues about 200 to 300, about 600 to 700 of SEQ ID NOs:2 and 5, or about 1100 to 1176 of SEQ ID NO:2 can be used to make an antibody against the phospholipid transporting region of the 67073 protein.

**[0170]** Antibodies reactive with, or specific or selective for, any of these regions, or other regions or domains described herein are provided.

**[0171]** Preferred epitopes encompassed by the antigenic peptide are regions of 67073 located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 67073 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 67073 protein and are thus likely to constitute surface residues useful for targeting antibody production.

**[0172]** In a preferred embodiment the antibody can bind to the extracellular portion of the 67073 protein, *e.g.*, it can bind to a whole cell which expresses the 67073 protein. In another embodiment, the antibody binds an intracellular portion of the 67073 protein.

**[0173]** In a preferred embodiment the antibody binds an epitope on any domain or region on 67073 proteins described herein.

**[0174]** Additionally, chimeric, humanized, and completely human antibodies are also within the scope of the invention. Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, *e.g.*, therapeutic treatment of human patients, and some diagnostic applications.

**[0175]** Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application

125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559.

5 [0176] A humanized or complementarity determining region (CDR)-grafted antibody will have at least one or two, but generally all three recipient CDR's (of heavy and or light immunoglobulin chains) replaced with a donor CDR. The antibody may be replaced with at least a portion of a non-human CDR or only some of the CDR's may be replaced with non-human CDR's. It is only necessary to replace the number of CDR's required for binding of the humanized antibody to a 67073 or a fragment thereof. Preferably, the donor will be a rodent antibody, *e.g.*, a rat or mouse antibody,

10 and the recipient will be a human framework or a human consensus framework. Typically, the immunoglobulin providing the CDR's is called the "donor" and the immunoglobulin providing the framework is called the "acceptor." In one embodiment, the donor immunoglobulin is a non-human (*e.g.*, rodent). The acceptor framework is a naturally-occurring (*e.g.*, a human) framework or a consensus framework, or a sequence about 85% or higher, preferably 90%, 95%, 99% or higher identical thereto.

15 [0177] As used herein, the term "consensus sequence" refers to the sequence formed from the most frequently occurring amino acids (or nucleotides) in a family of related sequences (See *e.g.*, Winnacker, (1987) *From Genes to Clones* (Verlagsgesellschaft, Weinheim, Germany). In a family of proteins, each position in the consensus sequence is occupied by the amino acid occurring most frequently at that position in the family. If two amino acids occur equally frequently, either can be included in the consensus sequence. A "consensus framework" refers to the framework region 20 in the consensus immunoglobulin sequence.

25 [0178] An antibody can be humanized by methods known in the art. Humanized antibodies can be generated by replacing sequences of the Fv variable region which are not directly involved in antigen binding with equivalent sequences from human Fv variable regions. General methods for generating humanized antibodies are provided by Morrison (1985) *Science* 229:1202-1207, by Oi *et al.* (1986) *BioTechniques* 4:214, and by Queen *et al.* US patent Nos. 5,585,089, 5,693,761 and 5,693,762, the contents of all of which are hereby incorporated by reference. Those methods include isolating, manipulating, and expressing the nucleic acid sequences that encode all or part of immunoglobulin Fv variable regions from at least one of a heavy or light chain. Sources of such nucleic acid are well known to those skilled in the art and, for example, may be obtained from a hybridoma producing an antibody against a 67073 polypeptide or fragment thereof. The recombinant DNA encoding the humanized antibody, or fragment thereof, can then be cloned into an appropriate expression vector.

30 [0179] Humanized or CDR-grafted antibodies can be produced by CDR-grafting or CDR substitution, wherein one, two, or all CDR's of an immunoglobulin chain can be replaced. See *e.g.*, U.S. Patent No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060; Winter US patent No. 5,225,539, the contents of all of which are hereby expressly incorporated by reference. Winter 35 describes a CDR-grafting method which may be used to prepare the humanized antibodies of the present invention (UK Patent Application GB 2188638A, filed on March 26, 1987; Winter US patent No. 5,225,539), the contents of which is expressly incorporated by reference.

40 [0180] Also within the scope of the invention are humanized antibodies in which specific amino acids have been substituted, deleted or added. Preferred humanized antibodies have amino acid substitutions in the framework region, such as to improve binding to the antigen. For example, a humanized antibody will have framework residues identical to the donor framework residue or to another amino acid other than the recipient framework residue. To generate such antibodies, a selected, small number of acceptor framework residues of the humanized immunoglobulin chain can be replaced by the corresponding donor amino acids. Preferred locations of the substitutions include amino acid residues adjacent to the CDR, or which are capable of interacting with a CDR (see *e.g.*, US patent No. 5,585,089). Criteria for 45 selecting amino acids from the donor are described in US 5,585,089, *e.g.*, columns 12-16 of US 5,585,089, the *e.g.*, columns 12-16 of US 5,585,089, the contents of which are hereby incorporated by reference. Other techniques for humanizing antibodies are described in Padlan *et al.* EP 519596 A1, published on December 23, 1992.

50 [0181] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93; and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

55 [0182] Completely human antibodies that recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described by Jespers *et al.* (1994) *BioTechnology* 12:899-903).

**[0183]** The anti-67073 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered as described in, for example, Colcher *et al.* (1999) *Ann. N Y Acad. Sci.* 880:263-80; and Reiter (1996) *Clin. Cancer Res.* 2:245-52. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 67073 protein.

**[0184]** In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it is an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, *e.g.*, it has a mutagenized or deleted Fc receptor binding region.

**[0185]** An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicine, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, maytansinoids, *e.g.*, maytansinol (see US Patent No. 5,208,020), CC-1065 (see US Patent Nos. 5,475,092, 5,585,499, 5,846,545) and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, CC-1065, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and antimitotic agents (*e.g.*, vincristine, vinblastine, taxol and maytansinoids). Radioactive ions include, but are not limited to iodine, yttrium and praseodymium.

**[0186]** The conjugates of the invention can be used for modifying a given biological response, the therapeutic moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the therapeutic moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, *pseudomonas* exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor,  $\square$ -interferon,  $\square$ -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

**[0187]** Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

**[0188]** An anti-67073 antibody (*e.g.*, monoclonal antibody) can be used to isolate 67073 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-67073 antibody can be used to detect 67073 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-67073 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance (*i.e.*, antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\square$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliflone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

**[0189]** In preferred embodiments, an antibody can be made by immunizing with a purified 67073 antigen, or a fragment thereof, *e.g.*, a fragment described herein, a membrane associated antigen, tissues, *e.g.*, crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions, *e.g.*, membrane fractions.

**[0190]** Antibodies which bind only a native 67073 protein, only denatured or otherwise non-native 67073 protein, or which bind both, are within the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes sometimes can be identified by identifying antibodies which bind to native but not denatured 67073 protein.

#### Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

**[0191]** In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, *e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses.

[0192] A vector can include a 67073 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 67073 proteins, mutant forms of 67073 proteins, fusion proteins, and the like).

[0193] The recombinant expression vectors of the invention can be designed for expression of 67073 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0194] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0195] Purified fusion proteins can be used in 67073 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific or selective for 67073 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks). To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0196] The 67073 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

[0197] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[0198] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\square$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[0199] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the

regulation of gene expression using antisense genes see Weintraub *et al.*, (1986) *Reviews - Trends in Genetics* 1:1.

**[0200]** Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, *e.g.*, a 67073 nucleic acid molecule within a recombinant expression vector or a 67073 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications can occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

**[0201]** A host cell can be any prokaryotic or eukaryotic cell. For example, a 67073 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or CV-1 origin, SV-40 (COS) cells). Other suitable host cells are known to those skilled in the art.

**[0202]** Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

**[0203]** A host cell of the invention can be used to produce (i.e., express) a 67073 protein. Accordingly, the invention further provides methods for producing a 67073 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 67073 protein has been introduced) in a suitable medium such that a 67073 protein is produced. In another embodiment, the method further includes isolating a 67073 protein from the medium or the host cell.

**[0204]** In another aspect, the invention features, a cell or purified preparation of cells which include a 67073 transgene, or which otherwise misexpress 67073. The cell preparation can consist of human or non-human cells, *e.g.*, rodent cells, *e.g.*, mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 67073 transgene, *e.g.*, a heterologous form of a 67073, *e.g.*, a gene derived from humans (in the case of a non-human cell). The 67073 transgene can be misexpressed, *e.g.*, overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpresses an endogenous 67073, *e.g.*, a gene the expression of which is disrupted, *e.g.*, a knockout. Such cells can serve as a model for studying disorders which are related to mutated or misexpressed 67073 alleles or for use in drug screening.

**[0205]** In another aspect, the invention features, a human cell, *e.g.*, a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 67073 polypeptide.

**[0206]** Also provided are cells, preferably human cells, *e.g.*, human hematopoietic or fibroblast cells, in which an endogenous 67073 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 67073 gene. The expression characteristics of an endogenous gene within a cell, *e.g.*, a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 67073 gene. For example, an endogenous 67073 gene which is "transcriptionally silent," *e.g.*, not normally expressed, or expressed only at very low levels, can be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, *e.g.*, Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991. Transgenic

#### 40 Animals

**[0207]** The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 67073 protein and for identifying and/or evaluating modulators of 67073 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, *e.g.*, a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, *e.g.*, a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 67073 gene has been altered by, *e.g.*, by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

**[0208]** Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 67073 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 67073 transgene in its genome and/or expression of 67073 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 67073 protein can further be bred to other transgenic animals carrying other transgenes.

[0209] 67073 proteins or polypeptides can be expressed in transgenic animals or plants, *e.g.*, a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, *e.g.*, a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

5 [0210] The invention also includes a population of cells from a transgenic animal, as discussed, *e.g.*, below. Uses

[0211] The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic).

10 [0212] The isolated nucleic acid molecules of the invention can be used, for example, to express a 67073 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect a 67073 mRNA (*e.g.*, in a biological sample) or a genetic alteration in a 67073 gene, and to modulate 67073 activity, as described further below. The 67073 proteins can be used to treat disorders characterized by insufficient or excessive production of a 67073 substrate or production of 67073 inhibitors. In addition, the 67073 proteins can be used to screen for naturally occurring 67073 substrates, to screen for drugs or compounds which modulate 67073 activity, as well as to treat disorders characterized by insufficient or excessive production of 67073 protein or production of 67073 protein forms which have decreased, aberrant or unwanted activity compared to 67073 wild type protein (*e.g.*, aberrant or deficient phospholipid transporter function or expression). Moreover, the anti-67073 antibodies of the invention can be used to detect and isolate 67073 proteins, regulate the bioavailability of 67073 proteins, and modulate 67073 activity.

20 [0213] A method of evaluating a compound for the ability to interact with, *e.g.*, bind, a subject 67073 polypeptide is provided. The method includes: contacting the compound with the subject 67073 polypeptide; and evaluating ability of the compound to interact with, *e.g.*, to bind or form a complex with the subject 67073 polypeptide. This method can be performed *in vitro*, *e.g.*, in a cell free system, or *in vivo*, *e.g.*, in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 67073 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 67073 polypeptide. Screening methods are discussed in more detail below.

#### Screening Assays:

30 [0214] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (*e.g.*, proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 67073 proteins, have a stimulatory or inhibitory effect on, for example, 67073 expression or 67073 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 67073 substrate. Compounds thus identified can be used to modulate the activity of target gene products (*e.g.*, 67073 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

35 [0215] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 67073 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 67073 protein or polypeptide or a biologically active portion thereof.

40 [0216] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, *e.g.*, Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

45 [0217] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909-13; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422-426; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678-85; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233-51.

50 [0218] Libraries of compounds can be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87: 6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra*).

**[0219]** In one embodiment, an assay is a cell-based assay in which a cell which expresses a 67073 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 67073 activity is determined. Determining the ability of the test compound to modulate 67073 activity can be accomplished by monitoring, for example, the ability of 67073 to reside within a membrane, hydrolyze ATP or transport a phospholipid.

**[0220]** The cell, for example, can be of mammalian origin, *e.g.*, human.

**[0221]** The ability of the test compound to modulate 67073 binding to a compound, *e.g.*, a 67073 substrate, or to bind to 67073 can also be evaluated. This can be accomplished, for example, by coupling the compound, *e.g.*, the substrate, with a radioisotope or enzymatic label such that binding of the compound, *e.g.*, the substrate, to 67073 can be determined by detecting the labeled compound, *e.g.*, substrate, in a complex. Alternatively, 67073 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 67073 binding to a 67073 substrate in a complex. For example, compounds (*e.g.*, 67073 substrates) can be labeled with  $^{125}\text{I}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$  or  $^{3}\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

**[0222]** The ability of a compound (*e.g.*, a 67073 substrate) to interact with 67073 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 67073 without the labeling of either the compound or the 67073. McConnell *et al.* (1992) *Science* 257: 1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 67073.

**[0223]** In yet another embodiment, a cell-free assay is provided in which a 67073 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 67073 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 67073 proteins to be used in assays of the present invention include fragments which participate in interactions with non-67073 molecules, *e.g.*, fragments with high surface probability scores.

**[0224]** Soluble and/or membrane-bound forms of isolated proteins (*e.g.*, 67073 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecylopoly(ethylene glycol ether)<sub>n</sub>, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

**[0225]** Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

**[0226]** The interaction between two molecules can also be detected, *e.g.*, using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule can simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label can be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

**[0227]** In another embodiment, determining the ability of the 67073 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, *e.g.*, Sjolander and Urbaniczky (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIACore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

**[0228]** In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not an-

chored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

**[0229]** It may be desirable to immobilize either 67073, an anti-67073 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 67073 protein, or interaction of a 67073 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/67073 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 67073 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 67073 binding or activity determined using standard techniques.

**[0230]** Other techniques for immobilizing either a 67073 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 67073 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

**[0231]** In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific or selective for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

**[0232]** In one embodiment, this assay is performed utilizing antibodies reactive with 67073 protein or target molecules but which do not interfere with binding of the 67073 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 67073 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 67073 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 67073 protein or target molecule.

**[0233]** Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas and Minton (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York.); and immunoprecipitation (see, for example, Ausubel *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard (1998) *J Mol Recognit* 11:141-8; Hage and Tweed (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer can also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

**[0234]** In a preferred embodiment, the assay includes contacting the 67073 protein or biologically active portion thereof with a known compound which binds 67073 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 67073 protein, wherein determining the ability of the test compound to interact with a 67073 protein includes determining the ability of the test compound to preferentially bind to 67073 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

**[0235]** The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 67073 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 67073 protein through modulation of the activity of a downstream effector of a 67073 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

**[0236]** To identify compounds that interfere with the interaction between the target gene product and its cellular or

extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

**[0237]** These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, *e.g.*, by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, *e.g.*, compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

**[0238]** In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (*e.g.*, a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific or selective for the species to be anchored can be used to anchor the species to the solid surface.

**[0239]** In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific or selective for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

**[0240]** Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; *e.g.*, using an immobilized antibody specific or selective for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific or selective for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

**[0241]** In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, *e.g.*, U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

**[0242]** In yet another aspect, the 67073 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *BioTechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 67073 ("67073-binding proteins" or "67073-bp") and are involved in 67073 activity. Such 67073-bps can be activators or inhibitors of signals by the 67073 proteins or 67073 targets as, for example, downstream elements of a 67073-mediated signaling pathway.

**[0243]** The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 67073 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an uni-

identified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the 67073 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 67073-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 67073 protein.

**[0244]** In another embodiment, modulators of 67073 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 67073 mRNA or protein evaluated relative to the level of expression of 67073 mRNA or protein in the absence of the candidate compound. When expression of 67073 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 67073 mRNA or protein expression. Alternatively, when expression of 67073 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 67073 mRNA or protein expression. The level of 67073 mRNA or protein expression can be determined by methods described herein for detecting 67073 mRNA or protein.

**[0245]** In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 67073 protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for aberrant or deficient phospholipid transporter function or expression.

**[0246]** This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 67073 modulating agent, an antisense 67073 nucleic acid molecule, a 67073-specific antibody, or a 67073-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

#### Detection Assays

**[0247]** Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 67073 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

#### 35 Chromosome Mapping

**[0248]** The 67073 nucleotide sequences or portions thereof can be used to map the location of the 67073 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 67073 sequences with genes associated with disease.

**[0249]** Briefly, 67073 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 67073 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 67073 sequences will yield an amplified fragment.

**[0250]** A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio *et al.* (1983) *Science* 220:919-924).

**[0251]** Other mapping strategies e.g., *in situ* hybridization (described in Fan *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 67073 to a chromosomal location.

**[0252]** Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York).

**[0253]** Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding se-

quences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

**[0254]** Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in McKusick, *Men-delian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature*, 325: 783-787.

**[0255]** Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 67073 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms. *Tissue Typing*

**[0256]** 67073 sequences can be used to identify individuals from biological samples using, *e.g.*, restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, *e.g.*, in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

**[0257]** Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 67073 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

**[0258]** Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 or SEQ ID NO:4 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 or SEQ ID NO:6 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

**[0259]** If a panel of reagents from 67073 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

#### Use of Partial 67073 Sequences in Forensic Biology

**[0260]** DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

**[0261]** The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 or SEQ ID NO: 4 (*e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 or SEQ ID NO: 4 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

**[0262]** The 67073 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 67073 probes can be used to identify tissue by species and/or by organ type.

**[0263]** In a similar fashion, these reagents, *e.g.*, 67073 primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

## Predictive Medicine

[0264] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[0265] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 67073.

[0266] Such disorders include, *e.g.*, a disorder associated with the misexpression of 67073 gene; a disorder of the endocrine system, nervous system, immune system or cardiovascular system.

[0267] The method includes one or more of the following:

[0268] detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 67073 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, *e.g.*, a mutation in the 5' control region;

[0269] detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 67073 gene;

[0270] detecting, in a tissue of the subject, the misexpression of the 67073 gene, at the mRNA level, *e.g.*, detecting a non-wild type level of an mRNA;

[0271] detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, *e.g.*, detecting a non-wild type level of a 67073 polypeptide.

[0272] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 67073 gene; an insertion of one or more nucleotides into the gene, a point mutation, *e.g.*, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, *e.g.*, a translocation, inversion, or deletion.

[0273] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, SEQ ID NO:4, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 67073 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, *e.g.*, *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[0274] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 67073 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 67073.

[0275] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[0276] In preferred embodiments the method includes determining the structure of a 67073 gene, an abnormal structure being indicative of risk for the disorder.

[0277] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 67073 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

## Diagnostic and Prognostic Assays

[0278] The presence, level, or absence of 67073 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 67073 protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes 67073 protein such that the presence of 67073 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 67073 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 67073 genes; measuring the amount of protein encoded by the 67073 genes; or measuring the activity of the protein encoded by the 67073 genes.

[0279] The level of mRNA corresponding to the 67073 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[0280] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 67073 nucleic acid, such as the nucleic acid of SEQ ID NO:1, SEQ ID NO:4, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 67073 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

**[0281]** In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 67073 genes.

**[0282]** The level of mRNA in a sample that is encoded by one of 67073 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (Kwoh *et al.*, (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta

10 Replicase (Lizardi *et al.*, (1988) *BioTechnology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

**[0283]** For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 67073 gene being analyzed.

**[0284]** In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 67073 mRNA, or genomic DNA, and comparing the presence of 67073 mRNA or genomic DNA in the control sample with the presence of 67073 mRNA or genomic DNA in the test sample.

**[0285]** A variety of methods can be used to determine the level of protein encoded by 67073. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

**[0286]** The detection methods can be used to detect 67073 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 67073 protein include enzyme linked immunosorbent assays (ELISAs), immuno-precipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 67073 protein include introducing into a subject a labeled anti-67073 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

**[0287]** In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 67073 protein, and comparing the presence of 67073 protein in the control sample with the presence of 67073 protein in the test sample.

**[0288]** The invention also includes kits for detecting the presence of 67073 in a biological sample. For example, the kit can include a compound or agent capable of detecting 67073 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 67073 protein or nucleic acid.

**[0289]** For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

**[0290]** For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

**[0291]** The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 67073 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

**[0292]** In one embodiment, a disease or disorder associated with aberrant or unwanted 67073 expression or activity

is identified. A test sample is obtained from a subject and 67073 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 67073 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 67073 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

**[0293]** The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 67073 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a phospholipid transporter disorder.

**[0294]** The methods of the invention can also be used to detect genetic alterations in a 67073 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 67073 protein activity or nucleic acid expression, such as a phospholipid transporter disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 67073-protein, or the mis-expression of the 67073 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 67073 gene; 2) an addition of one or more nucleotides to a 67073 gene; 3) a substitution of one or more nucleotides of a 67073 gene, 4) a chromosomal rearrangement of a 67073 gene; 5) an alteration in the level of a messenger RNA transcript of a 67073 gene, 6) aberrant modification of a 67073 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 67073 gene, 8) a non-wild type level of a 67073-protein, 9) allelic loss of a 67073 gene, and 10) inappropriate post-translational modification of a 67073-protein.

**[0295]** An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 67073-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 67073 gene under conditions such that hybridization and amplification of the 67073 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

**[0296]** In another embodiment, mutations in a 67073 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

**[0297]** In other embodiments, genetic mutations in 67073 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 67073 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

**[0298]** In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 67073 gene and detect mutations by comparing the sequence of the sample 67073 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.* (1995) *Biotechniques* 19:448-53), including sequencing by mass spectrometry.

**[0299]** Other methods for detecting mutations in the 67073 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

**[0300]** In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize

mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 67073 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

**[0301]** In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 67073 genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 67073 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

**[0302]** In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

**[0303]** Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci USA* 86:6230).

**[0304]** Alternatively, allele specific amplification technology which depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification can also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189-193). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

**[0305]** The methods described herein can be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 67073 gene.

#### Use of 67073 Molecules as Surrogate Markers

**[0306]** The 67073 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 67073 molecules of the invention can be detected, and can be correlated with one or more biological states *in vivo*. For example, the 67073 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

**[0307]** The 67073 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "phar-

macodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a biological

5 tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of  
10 drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (e.g., a 67073 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-67073 antibodies can be  
15 employed in an immune-based detection system for a 67073 protein marker, or 67073-specific radiolabeled probes can be used to detect a 67073 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

20 [0308] The 67073 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers  
25 in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 67073 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 67073 DNA can correlate with a 67073 drug response. The use of pharmacogenomic markers  
30 therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy. Pharmaceutical Compositions

35 [0309] The nucleic acid and polypeptides, fragments thereof, as well as anti-67073 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used  
herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

40 [0310] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the  
45 adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

50 [0311] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should  
55 be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the

like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

**[0312]** Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

**[0313]** Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition.

The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

**[0314]** For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

**[0315]** Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

**[0316]** The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

**[0317]** In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

**[0318]** It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

**[0319]** Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD<sub>50</sub> (the dose lethal to 50% of the population) and the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD<sub>50</sub> / ED<sub>50</sub>. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

**[0320]** The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC<sub>50</sub> (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

**[0321]** As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) rang-

es from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody, unconjugated or conjugated as described herein, can include a single treatment or, preferably, can include a series of treatments.

**[0322]** For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

**[0323]** The present invention encompasses agents which modulate expression or activity. An agent can, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

**[0324]** Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

**[0325]** The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U. S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

**[0326]** The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

#### 45 Methods of Treatment:

**[0327]** The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 67073 expression or activity. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

**[0328]** With regards to both prophylactic and therapeutic methods of treatment, such treatments can be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the

study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 67073 molecules of the present invention or 67073 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

**[0329]** In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 67073 expression or activity, by administering to the subject a 67073 or an agent which modulates 67073 expression or at least one 67073 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 67073 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 67073 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 67073 aberrance, for example, a 67073, 67073 agonist or 67073 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

**[0330]** It is possible that some 67073 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

**[0331]** The 67073 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of blood coagulation disorders; apoptotic disorders, immune (e.g. inflammatory) disorders; hormonal disorders; disorders involving aberrant vesicle function and/or formation, such as neurological disorders (e.g. brain disorders); disorders of inter- or intra-cellular communication; disorders of tissue function, such as cardiac function or musculoskeletal function; cellular proliferation, growth, or differentiation disorders, viral diseases, and testicular disorders, all of which are described above. The molecules of the invention also can act as novel diagnostic targets and therapeutic agents for controlling one or more of disorders associated with bone metabolism, liver disorders, pain disorders or metabolic disorders as described below.

**[0332]** Aberrant expression and/or activity of 67073 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 67073 molecules in bone cells, e.g. osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, 67073 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 67073 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

**[0333]** Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of pre-existing fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, Al-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein can be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

**[0334]** Additionally, 67073 can play an important role in the regulation of metabolism or pain disorders. Diseases of

metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H. L. (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

**[0335]** As discussed, successful treatment of 67073 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 67073 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, human, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')<sub>2</sub> and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

**[0336]** Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

**[0337]** It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

**[0338]** Another method by which nucleic acid molecules can be utilized in treating or preventing a disease characterized by 67073 expression is through the use of aptamer molecules specific for 67073 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically or selectively bind to protein ligands (see, e.g., Osborne *et al.* (1997) *Curr. Opin. Chem Biol.* 1: 5-9; and Patel (1997) *Curr Opin Chem Biol* 1:32-46). Since nucleic acid molecules can in many cases be more conveniently introduced into target cells than therapeutic protein molecules can be, aptamers offer a method by which 67073 protein activity can be specifically decreased without the introduction of drugs or other molecules which can have pluripotent effects.

**[0339]** Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies can, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of 67073 disorders. For a description of antibodies, see the Antibody section above.

**[0340]** In circumstances wherein injection of an animal or a human subject with a 67073 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 67073 through the use of anti-idiotypic antibodies (see, for example, Herlyn (1999) *Ann Med* 31:66-78; and Bhattacharya-Chatterjee and Foon (1998) *Cancer Treat Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 67073 protein.

**[0341]** Vaccines directed to a disease characterized by 67073 expression can also be generated in this fashion.

**[0341]** In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies can be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

**[0342]** The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 67073 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures as described above.

**[0343]** The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models

to achieve a circulating plasma concentration range that includes the  $IC_{50}$  (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

5 [0344] Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays can utilize antibody mimics and/or "bio-sensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 67073 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix

10 which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al* (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis *et al* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 67073 can be readily monitored and used in calculations of  $IC_{50}$ .

15 [0345] Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly 20 optimized based on its individual  $IC_{50}$ . An rudimentary example of such a "biosensor" is discussed in Kriz *et al* (1995) *Analytical Chemistry* 67:2142-2144.

25 [0346] Another aspect of the invention pertains to methods of modulating 67073 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 67073 or agent that modulates one or more of the activities of 67073 protein activity associated with the cell. An agent that modulates 67073 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 67073 protein (e.g., a 67073 substrate or receptor), a 67073 antibody, a 67073 agonist or antagonist, a peptidomimetic of a 67073 agonist or antagonist, or other small molecule.

30 [0347] In one embodiment, the agent stimulates one or 67073 activities. Examples of such stimulatory agents include active 67073 protein and a nucleic acid molecule encoding 67073. In another embodiment, the agent inhibits one or more 35 67073 activities. Examples of such inhibitory agents include antisense 67073 nucleic acid molecules, anti-67073 antibodies, and 67073 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 67073 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up regulates or down regulates) 67073 expression or activity. In another embodiment, the method involves administering a 67073 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 67073 expression or activity.

40 [0348] Stimulation of 67073 activity is desirable in situations in which 67073 is abnormally downregulated and/or in which increased 67073 activity is likely to have a beneficial effect. For example, stimulation of 67073 activity is desirable in situations in which a 67073 is downregulated and/or in which increased 67073 activity is likely to have a beneficial effect. Likewise, inhibition of 67073 activity is desirable in situations in which 67073 is abnormally upregulated and/or in which decreased 67073 activity is likely to have a beneficial effect.

45 Pharmacogenomics

50 [0349] The 67073 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 67073 activity (e.g., 67073 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 67073-associated disorders (e.g., aberrant or deficient phospholipid transporter function or expression) associated with aberrant or unwanted 67073 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) can be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician can consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 67073 molecule or 67073 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 67073 molecule or 67073 modulator.

55 [0350] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum *et al*. (1996) *Clin. Exp.*

*Pharmacol. Physiol.* 23:983-985 and Linder *et al.* (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (antimalarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

**[0351]** One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP can occur once per every 1000 bases of DNA. A SNP can be involved in a disease process, however, the vast majority can not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that can be common among such genetically similar individuals.

**[0352]** Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 67073 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

**[0353]** Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 67073 molecule or 67073 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

**[0354]** Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 67073 molecule or 67073 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

**[0355]** The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 67073 genes of the present invention, wherein these products can be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 67073 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells, will become sensitive to treatment with an agent to which the unmodified target cells were resistant.

**[0356]** Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 67073 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 67073 gene expression, protein levels, or upregulate 67073 activity, can be monitored in clinical trials of subjects exhibiting decreased 67073 gene expression, protein levels, or downregulated 67073 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 67073 gene expression, protein levels, or downregulate 67073 activity, can be monitored in clinical trials of subjects exhibiting increased 67073 gene expression, protein levels, or upregulated 67073 activity. In such clinical trials, the expression or activity of a 67073 gene, and preferably, other genes that have been implicated in, for example, a phospholipid transporter-associated or another 67073-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

#### 50 Other Embodiments

**[0357]** In another aspect, the invention features a method of analyzing a plurality of capture probes. The method is useful, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence, wherein the capture probes are from a cell or subject which expresses 67073 or from a cell or subject in which a 67073 mediated response has been elicited; contacting the array with a 67073 nucleic acid (preferably purified), a 67073 polypeptide

(preferably purified), or an anti-67073 antibody, and thereby evaluating the plurality of capture probes. Binding, *e.g.*, in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, *e.g.*, by a signal generated from a label attached to the 67073 nucleic acid, polypeptide, or antibody.

**[0358]** The capture probes can be a set of nucleic acids from a selected sample, *e.g.*, a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

**[0359]** The method can include contacting the 67073 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, *e.g.*, to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, *e.g.*, a wild type, normal, or non-diseased, non-stimulated, sample, *e.g.*, a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, *e.g.*, a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, *e.g.*, a biological fluid, tissue, or cell sample.

**[0360]** The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 67073. Such methods can be used to diagnose a subject, *e.g.*, to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

**[0361]** The method can be used to detect SNPs, as described above.

**[0362]** In another aspect, the invention features, a method of analyzing 67073, *e.g.*, analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 67073 nucleic acid or amino acid sequence; comparing the 67073 sequence with one or more preferably a plurality of sequences from a collection of sequences, *e.g.*, a nucleic acid or protein sequence database; to thereby analyze 67073.

**[0363]** The method can include evaluating the sequence identity between a 67073 sequence and a database sequence. The method can be performed by accessing the database at a second site, *e.g.*, over the internet. Preferred databases include CenBank™ and SwissProt.

**[0364]** In another aspect, the invention features, a set of oligonucleotides, useful, *e.g.*, for identifying SNP's, or identifying specific alleles of 67073. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, *e.g.*, an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

**[0365]** The sequences of 67073 molecules are provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 67073 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, *e.g.*, an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exist in nature or in purified form.

**[0366]** A 67073 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media.

**[0367]** The medium is adapted or configured for having thereon 67073 sequence information of the present invention.

**[0368]** As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

**[0369]** As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 67073 sequence information.

**[0370]** A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a 67073 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

**[0370]** By providing the 67073 nucleotide or amino acid sequences of the invention in computer readable form, the

skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

**[0371]** The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder, wherein the method comprises the steps of determining 67073 sequence information associated with the subject and based on the 67073 sequence information, determining whether the subject has a phospholipid transporter-associated or another 67073-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

**[0372]** The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a disease associated with 67073, wherein the method comprises the steps of determining 67073 sequence information associated with the subject, and based on the 67073 sequence information, determining whether the subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

**[0373]** The present invention also provides in a network, a method for determining whether a subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder, said method comprising the steps of receiving 67073 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 67073 and/or corresponding to a phospholipid transporter-associated or another 67073-associated disease or disorder, and based on one or more of the phenotypic information, the 67073 information (e.g., sequence information and/or information related thereto), and the acquired information, determining whether the subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

**[0374]** The present invention also provides a business method for determining whether a subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder, said method comprising the steps of receiving information related to 67073 (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 67073 and/or related to a phospholipid transporter-associated or another 67073-associated disease or disorder, and based on one or more of the phenotypic information, the 67073 information, and the acquired information, determining whether the subject has a phospholipid transporter-associated or another 67073-associated disease or disorder or a pre-disposition to a phospholipid transporter-associated or another 67073-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

**[0375]** The invention also includes an array comprising a 67073 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 67073. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

**[0376]** In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue if ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at

the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

**[0377]** In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a phospholipid transporter-associated or another 67073-associated disease or disorder, progression of phospholipid transporter-associated or another 67073-associated disease or disorder, and processes, such a cellular transformation associated with the phospholipid transporter-associated or another 67073-associated disease or disorder.

**[0378]** The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (e.g., ascertaining the effect of 67073 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

**[0379]** The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (e.g., including 67073) that could serve as a molecular target for diagnosis or therapeutic intervention.

**[0380]** As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

**[0381]** Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

**[0382]** Thus, the invention features a method of making a computer readable record of a sequence of a 67073 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

**[0383]** In another aspect, the invention features a method of analyzing a sequence. The method includes: providing a 67073 sequence, or record, in computer readable form; comparing a second sequence to the 67073 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 67073 sequence includes a sequence being compared. In a preferred embodiment the 67073 or second sequence is stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. E.g., the 67073 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

**[0384]** The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

**45** Equivalents

**[0385]** Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein.

EP 1 225 182 A2

SEQUENCE LISTING

5 <110> Curtis, Rory A.J.  
Millennium Pharmaceuticals Inc.

10 <120> 67073, A Human Phospholipid Transporter  
Family Member and Uses Therefor

15 <130> MPI2001-015P2R (M)

20 <150> Our Ref: P71870EP  
<151> 2001-01-17

25 <150> 60/289, 358  
<151> 2001-05-08

30 <160> 11

35 <170> FastSEQ for Windows Version 4.0

40 <210> 1  
<211> 3878  
<212> DNA  
<213> Homo sapiens

45 <220>  
<221> CDS  
<222> (20) ... (3550)

50 <400> 1  
30 cctggacaaa gctcttaag atg tcc ctg ccg cgg agg tcg agg atc cgc tcg 52  
Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser  
1 5 10

35 tcc gtg gga cct gtt cgt tct tct ttg ggc tat aag aag gca gat gat 100  
Ser Val Gly Pro Val Arg Ser Ser Leu Gly Tyr Lys Lys Ala Asp Asp  
15 20 25

40 gag atg tcc ccg gcc acg tct gtt gga gac cag ctg gag gca ccc gcc 148  
Glu Met Ser Arg Ala Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala  
30 35 40

45 cgc acc att tac ctc aac caa ccg cat ctc aac aaa ttc cgc gac aac 196  
Arg Thr Ile Tyr Leu Asn Gln Prc His Leu Asn Lys Phe Arg Asp Asn  
45 50 55

50 cag atc agt acg gcc aag tac agc gtg ttg aca ttt cta cct cga ttc 244  
Gln Ile Ser Thr Ala Lys Tyr Ser Val Leu Thr Phe Leu Pro Arg Phe  
60 65 70 75

55 ttg tat gag cag att aga aga gct gct aat gcc ttc ttt ctc ttc att 292  
Leu Tyr Glu Gln Ile Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile  
80 85 90

60 gcc tta tta cag caa att cca gat gta tct cca aca gga aga tat acc 340  
Ala Leu Leu Gln Gln Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr  
95 100 105

65 acc ctg gtg cca ttg atc att att tta aca att gca ggc atc aag gag 388

EP 1 225 182 A2

	Thr Leu Val Pro Leu Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu		
	110	115	120
5	att gta gaa gat ttt aag cga cac aag gca gac aat gca gtt aac aaa Ile Val Glu Asp Phe Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys	436	
	125	130	135
10	aag aaa aca ata gtg tta aga aat ggt atg tgg cat acc att atg tgg Lys Lys Thr Ile Val Leu Arg Asn Gly Met Trp His Thr Ile Met Trp	484	
	140	145	150
	155		
15	aaa gag gtg gca gtg gga gac att gtg aag gtc gtc aat ggg cag tat Lys Glu Val Ala Val Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr	532	
	160	165	170
20	ctt cca gca gat gtg gtc ctg ctg tca tcc agt gaa cct cag gca atg Leu Pro Ala Asp Val Val Leu Leu Ser Ser Ser Glu Pro Gln Ala Met	580	
	175	180	185
25	tgt tat gtt gaa aca gct aat ctg gat ggg gag acg aac ctt aaa ata Cys Tyr Val Glu Thr Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile	628	
	190	195	200
30	cgt cag ggt ttg agt cac act gct gac atg caa aca cgt gaa gtt ctg Arg Gln Gly Leu Ser His Thr Ala Asp Met Gln Thr Arg Glu Val Leu	676	
	205	210	215
35	atg aag tta tct gga act ata gag ttt gaa ggg ccc aac cgc cac ctc Met Lys Leu Ser Gly Thr Ile Glu Cys Glu Gly Pro Asn Arg His Leu	724	
	220	225	230
	235		
40	tat gac ttc act gga aac ttg aac tta gat ggg aaa agc ctt gtt gcc Tyr Asp Phe Thr Gly Asn Leu Asn Leu Asp Gly Lys Ser Leu Val Ala	772	
	240	245	250
45	ctt ggg cct gac cag atc tta tta aga ggt aca cag ctt aga aat act Leu Gly Pro Asp Gln Ile Leu Leu Arg Gly Thr Gln Leu Arg Asn Thr	820	
	255	260	265
50	cag tgg gtc ttt ggc ata gtt gtt tat act gga cac gac acc aaa ctc Gln Trp Val Phe Gly Ile Val Val Tyr Thr Gly His Asp Thr Lys Leu	868	
	270	275	280
55	atg cag aat tca acc aaa gcg cct ctc aag aga tca aat gtt gag aag Met Gln Asn Ser Thr Lys Ala Pro Leu Lys Arg Ser Asn Val Glu Lys	916	
	285	290	295
60	gtg act aac gtg cag atc ctg gtg ttg ttt ggc atc ctc ttg gtc atg Val Thr Asn Val Gln Ile Leu Val Leu Phe Gly Ile Leu Leu Val Met	964	
	300	305	310
	315		
65	gcc ttg gtg agc tcg gcg ggg ggc ctg tac tgg aac agg tct cat ggt Ala Leu Val Ser Ser Ala Gly Ala Leu Tyr Trp Asn Arg Ser His Gly	1012	
	320	325	330
70	gaa aag aac tgg tac atc aag aag atg gac acc acc tca gat aat ttt Glu Lys Asn Trp Tyr Ile Lys Lys Met Asp Thr Thr Ser Asp Asn Phe	1060	
	335	340	345
75	gga tac aac cta ctg acg ttc atc atc tta tac aac aat ctt att ccc Gly Tyr Asn Leu Leu Thr Phe Ile Ile Leu Tyr Asn Asn Leu Ile Pro	1108	

EP 1 225 182 A2

	350	355	360	
5	atc agt ctg ttg gtg ast ctt gag gtt gtg aag tat act caa gcc ctt Ile Ser Leu Leu Val Thr Leu Glu Val Val Lys Tyr Thr Gln Ala Leu 365 370 375			1156
10	ttc ata aac tgg gac aca gat atg tat tat ata gga aat gac act cct Phe Ile Asn Trp Asp Thr Asp Met Tyr Tyr Ile Gly Asn Asp Thr Pro 380 385 390 395			1204
15	gcc atg gcc agg aca tca aac ctt aat gaa gag ctt ggg cag gtg aaa Ala Met Ala Arg Thr Ser Asn Leu Asn Glu Glu Leu Gly Gln Val Lys 400 405 410			1252
20	tat ctc ttt tct gac aag act gga acg ctt aca tgc aat atc atg aac Tyr Leu Phe Ser Asp Lys Thr Gly Thr Leu Thr Cys Asn Ile Met Asn 415 420 425			1300
25	ttt aag aag tgc agc att gcc gga gta acc tat ggt cac ttc cca gaa Phe Lys Lys Cys Ser Ile Ala Gly Val Thr Tyr Gly His Phe Pro Glu 430 435 440			1348
30	tgc gca aga gag ccc tct tca gat gac ttc tgt cgg atg cct ccc Leu Ala Arg Glu Pro Ser Ser Asp Asp Phe Cys Arg Met Pro Pro Pro 445 450 455			1396
35	tgt agt gat tcc tgt gac ttt gat gac ccc agg ctg ttg aag aac att Cys Ser Asp Ser Cys Asp Phe Asp Asp Pro Arg Leu Leu Lys Asn Ile 460 465 470 475			1444
40	gag gat cgc cat ccc aca gcc cct tgc att cag gag ttc ctc acc ctt Glu Asp Arg His Pro Thr Ala Pro Cys Ile Gln Glu Phe Leu Thr Leu 480 485 490			1492
45	ctg gcc gtg tgc cac acg gtt gtt cct gag aag gat gga gat aac atc Leu Ala Val Cys His Thr Val Val Pro Glu Lys Asp Gly Asp Asn Ile 495 500 505			1540
50	atc tac cag gcc tct tcc cca gat gaa gct gct ttg gtg aaa gga gct Ile Tyr Gln Ala Ser Ser Pro Asp Glu Ala Ala Leu Val Lys Gly Ala 510 515 520			1588
55	aaa aag ctg ggc ttt gtc ttc aca gcc aga aca cca ttc tca gtc atc Lys Lys Leu Gly Phe Val Phe Thr Ala Arg Thr Pro Phe Ser Val Ile 525 530 535			1636
60	ata gaa gcg atg gga cag gaa caa aca ttc gga atc ctt aat gtc ctg Ile Glu Ala Met Gly Gln Glu Gln Thr Phe Gly Ile Leu Asn Val Leu 540 545 550 555			1684
65	gaa ttt tct agt gac aga aaa aga atg tct gta att gtt cga act cct Glu Phe Ser Ser Asp Arg Lys Arg Met Ser Val Ile Val Arg Thr Pro 560 565 570			1732
70	tca gga cga ctt cgg ctt tac tgt aaa ggg gct gat aat gtg att ttt Ser Gly Arg Leu Arg Leu Tyr Cys Lys Gly Ala Asp Asn Val Ile Phe 575 580 585			1780
75	gag aga ctt tca aaa gac tca aaa tat atg gag gaa aca tta tgc cat Glu Arg Leu Ser Lys Asp Ser Lys Tyr Met Glu Glu Thr Leu Cys His 590 595 600			1828

EP 1 225 182 A2

	ctg gaa tac ttt gcc acg gaa ggc ttg cgg act ctc tgt gtg gct tat Leu Glu Tyr Phe Ala Thr Glu Gly Leu Arg Thr Leu Cys Val Ala Tyr 605 610 615	1876
5	gct gat ctc tct gag aat gag tat gag gag tgg ctg aaa gtc tat cag Ala Asp Leu Ser Glu Asn Glu Tyr Glu Glu Trp Leu Lys Val Tyr Gln 620 625 630 635	1924
10	gaa gcc agc acc ata ttg aag gac aga gct caa cgg ttg gaa gag tgt Glu Ala Ser Thr Ile Leu Lys Asp Arg Ala Gln Arg Leu Glu Glu Cys 640 645 650	1972
15	tac gag atc att gag aag aat ttg ctg cta ctt gga gcc aca gcc ata Tyr Glu Ile Ile Glu Lys Asn Leu Leu Leu Gly Ala Thr Ala Ile 655 660 665	2020
20	gaa gat cgc ctt caa gca gga gtt cca gaa acc atc gca aca ctg ttg Glu Asp Arg Leu Gln Ala Gly Val Pro Glu Thr Ile Ala Thr Leu Leu 670 675 680	2068
25	aag gca gaa att aaa ata tgg gtg ttg aca gga gac aaa caa gaa act Lys Ala Glu Ile Lys Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr 685 690 695	2116
30	gcg att aat ata ggg tat tcc tgc cga ttg gta tcg cag aat atg gcc Ala Ile Asn Ile Gly Tyr Ser Cys Arg Leu Val Ser Gln Asn Met Ala 700 705 710 715	2164
35	ctt atc cta ttg aag gag gac tct ttg gat gcc aca agg gca gcc att Leu Ile Leu Leu Lys Glu Asp Ser Leu Asp Ala Thr Arg Ala Ala Ile 720 725 730	2212
40	act cag cac tgc act gac ctt ggg aat ttg ctg ggc aag gaa aat gac Thr Gln His Cys Thr Asp Leu Gly Asn Leu Leu Gly Lys Glu Asn Asp 735 740 745	2260
45	gtg gcc ctc atc atc gat ggc cac acc ctg aag tac gcg ctc tcc ttc Val Ala Leu Ile Ile Asp Gly His Thr Leu Lys Tyr Ala Leu Ser Phe 750 755 760	2308
50	gaa gtc cgg agg agt ttc ctg gat ttg gca ctc tcg tgc aaa gcg gtc Glu Val Arg Arg Ser Phe Leu Asp Leu Ala Leu Ser Cys Lys Ala Val 765 770 775	2356
55	ata tgc tgc aga gtg tct cct ctg cag aag tct gag ata gtg gat gtg Ile Cys Cys Arg Val Ser Pro Leu Gln Lys Ser Glu Ile Val Asp Val 780 785 790 795	2404
	gtg aag aag cgg gtg aag gcc atc acc ctc gcc atc gga gac ggc gcc Val Lys Lys Arg Val Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala 800 805 810	2452
	aac gat gtc ggg atg atc cag aca gcc cac gtg ggt gtg gga atc agt Asn Asp Val Gly Met Ile Gln Thr Ala His Val Gly Val Gly Ile Ser 815 820 825	2500
	ggg aat gaa ggc atg cag gcc acc aac aac tcg gat tac gcc atc gca Gly Asn Glu Gly Met Gln Ala Thr Asn Asn Ser Asp Tyr Ala Ile Ala 830 835 840	2548

EP 1 225 182 A2

5	cag ttt tcc tac tta gag aag ctt ctg ttg gtt cat gga gcc tgg agc Gln Phe Ser Tyr Leu Glu Lys Leu Leu Val His Gly Ala Trp Ser 845 850 855	2596
10	tac aac cgg gtg acc aag tgc atc ttg tac tgc ttc tat aag aac gtg Tyr Asn Arg Val Thr Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Val 860 865 870 875	2644
15	gtc ctg tat att att gag ctt tgg ttc gcc ttt gtt aat gga ttt tct Val Leu Tyr Ile Ile Glu Leu Trp Phe Ala Phe Val Asn Gly Phe Ser 880 885 890	2692
20	ggg cag att tta ttt gaa cgt tgg tgc atc ggc ctg tac aat gtg att Gly Gln Ile Leu Phe Glu Arg Trp Cys Ile Gly Leu Tyr Asn Val Ile 895 900 905	2740
25	ttc acc gct ttg ccg ccc ttc act ctg gga atc ttt gag agg tct tgc Phe Thr Ala Leu Pro Pro Phe Thr Leu Gly Ile Phe Glu Arg Ser Cys 910 915 920	2788
30	act cag gag agc atg ctc agg ttt ccc cag ctc tac aaa atc acc cag Thr Gln Glu Ser Met Leu Arg Phe Pro Gln Leu Tyr Lys Ile Thr Gln 925 930 935	2836
35	aat ggc gaa ggc ttc aac aca aag gtt ttc tgg ggt cac tgc atc aac Asn Gly Glu Gly Phe Asn Thr Lys Val Phe Trp Gly His Cys Ile Asn 940 945 950 955	2884
40	gcc ttg gtc cac tcc ctc atc ctc ttc tgg ttt ccc atg aaa gct ctg Ala Leu Val His Ser Leu Ile Leu Phe Trp Phe Pro Met Lys Ala Leu 960 965 970	2932
45	gag cat gat act gtg ttg aca agt ggt cat gct acc gac tat tta ttt Glu His Asp Thr Val Leu Thr Ser Gly His Ala Thr Asp Tyr Leu Phe 975 980 985	2980
50	gtt gga aat att gtt tac aca tat gtt gtt act gtt tgt ctg aaa Val Gly Asn Ile Val Tyr Thr Tyr Val Val Val Thr Val Cys Leu Lys 990 995 1000	3028
55	gct ggt ttg gag acc aca gct tgg act aaa ttc agt cat ctg gct gtc Ala Gly Leu Glu Thr Thr Ala Trp Thr Lys Phe Ser His Leu Ala Val 1005 1010 1015	3076
60	tgg gga agc atg ctc acc tgg ctg gtg ttt ttt ggc atc tac tcg acc Trp Gly Ser Met Leu Thr Trp Leu Val Phe Phe Gly Ile Tyr Ser Thr 1020 1025 1030 1035	3124
65	atc tgg ccc acc att ccc att gct cca gat atg aga gga cag gca act Ile Trp Pro Thr Ile Pro Ile Ala Pro Asp Met Arg Gly Gln Ala Thr 1040 1045 1050	3172
70	atg gtc ctg agc tcc gca cac ttc tgg ttg gga tta ttt ctg gtt cct Met Val Leu Ser Ser Ala His Phe Trp Leu Gly Leu Phe Leu Val Pro 1055 1060 1065	3220
75	act gcc tgt ttg att gaa gat gtg gca tgg aga gca gcc aag cac acc Thr Ala Cys Leu Ile Glu Asp Val Ala Trp Arg Ala Ala Lys His Thr 1070 1075 1080	3268
80	tgc aaa aag aca ttg ctg gag gag gtg cag gag ctg gaa acc aag tct	3316

EP 1 225 182 A2

	Cys Lys Lys Thr Leu Leu Glu Glu Val Gln Glu Leu Glu Thr Lys Ser	
	1085 1090 1095	
5	cga gtc ctg gga aaa gcg gtg ctg cgg gat agc aat gga aag agg ctg Arg Val Leu Gly Lys Ala Val Leu Arg Asp Ser Asn Gly Lys Arg Leu	3364
	1100 1105 1110 1115	
10	aac gag cgc gac cgc ctg atc aag agg ctg ggc cgg aag acg ccc cgc Asn Glu Arg Asp Arg Leu Ile Lys Arg Leu Gly Arg Lys Thr Pro Pro	3412
	1120 1125 1130	
	acg ctg ttc cgg ggc agc tcc ctg cag cag ggc gtc ccg cat ggg tat Thr Leu Phe Arg Gly Ser Ser Leu Gln Gln Gly Val Pro His Gly Tyr	3460
	1135 1140 1145	
15	gct ttt tct caa gaa gaa cac gga gct gtt agt cag gaa gaa gtc atc Ala Phe Ser Gln Glu His Gly Ala Val Ser Gln Glu Glu Val Ile	3508
	1150 1155 1160	
20	cgt gct tat gac acc acc aaa aag aaa tcc agg aag aaa taa Arg Ala Tyr Asp Thr Thr Lys Lys Ser Arg Lys Lys *	3550
	1165 1170 1175	
	gacatgaatt ttccctgactg atcttaggaa agagattcag ttgtttgcac ccagtgttaa 3610 cacatctttg tcagagaaga ctggcgctcg cagccaaaac accagggaaac acatttctgt 3670 ggccttagcc aagcagtttg ttagttacat attccctcgcc 3730 aggggaagct atctttgccc tcccaactcg tctgcagtgc ttagcctaac ttttgtttat 3790 gtcgttatga agcattcaac tgcgtctgt gaggtgtgaa attaaaaaca ttatgtttca 3850 ccaaaaaaaaaaaaaaaagg gccggccgc 3878	
25		
30	<210> 2 <211> 1176 <212> PRT <213> Homo sapiens	
	<400> 2	
35	Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser Ser Val Gly Pro Val 1 5 10 15 Arg Ser Ser Leu Gly Tyr Lys Ala Asp Asp Glu Met Ser Arg Ala 20 25 30 Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala Arg Thr Ile Tyr Leu 35 40 45	
40	Asn Gln Pro His Leu Asn Lys Phe Arg Asp Asn Gln Ile Ser Thr Ala 50 55 60 Lys Tyr Ser Val Leu Thr Phe Leu Pro Arg Phe Leu Tyr Glu Gln Ile 65 70 75 80 Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile Ala Leu Leu Gln Gln 85 90 95	
45	Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr Thr Leu Val Pro Leu 100 105 110 Ile Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu Ile Val Glu Asp Phe 115 120 125 Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys Lys Lys Thr Ile Val 130 135 140	
50	Leu Arg Asn Gly Met Trp His Thr Ile Met Trp Lys Glu Val Ala Val 145 150 155 160 Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr Leu Pro Ala Asp Val 165 170 175 Val Leu Leu Ser Ser Glu Pro Gln Ala Met Cys Tyr Val Glu Thr 180 185 190 Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile Arg Gln Gly Leu Ser	

EP 1 225 182 A2

	195	200	205
	His Thr Ala Asp Met Gln	Thr Arg Glu Val Leu Met	Lys Leu Ser Gly
	210	215	220
5	Thr Ile Glu Cys Glu Gly	Pro Asn Arg His Leu Tyr	Asp Phe Thr Gly
	225	230	235
	Asn Leu Asn Leu Asp Gly	Ser Leu Val Ala Leu	Gly Pro Asp Gln
	245	250	255
	Ile Leu Leu Arg Gly	Thr Gln Leu Arg Asn	Thr Gln Trp Val Phe Gly
10	260	265	270
	Ile Val Val Tyr Thr Gly	His Asp Thr Lys Leu Met	Gln Asn Ser Thr
	275	280	285
	Lys Ala Pro Leu Lys Arg	Ser Asn Val Glu Lys	Val Thr Asn Val Gln
	290	295	300
15	Ile Leu Val Leu Phe Gly	Ile Leu Leu Val Met Ala	Leu Val Ser Ser
	305	310	315
	Ala Gly Ala Leu Tyr Trp	Asn Arg Ser His Gly	Glu Lys Asn Trp Tyr
	325	330	335
	Ile Lys Lys Met Asp Thr	Thr Ser Asp Asn Phe	Gly Tyr Asn Leu Leu
	340	345	350
20	Thr Phe Ile Ile Leu Tyr	Asn Asn Leu Ile Pro	Ile Ser Leu Leu Val
	355	360	365
	Thr Leu Glu Val Val Lys	Tyr Thr Gln Ala Leu Phe	Ile Asn Trp Asp
	370	375	380
	Thr Asp Met Tyr Tyr	Ile Gly Asn Asp Thr	Pro Ala Met Ala Arg
	385	390	395
25	Ser Asn Leu Asn Glu Glu	Leu Gly Gln Val Lys	Tyr Leu Phe Ser Asp
	405	410	415
	Lys Thr Gly Thr Leu Thr Cys	Asn Ile Met Asn Phe	Lys Lys Cys Ser
	420	425	430
	Ile Ala Gly Val Thr Tyr	Gly His Phe Pro	Glu Leu Ala Arg Glu Pro
	435	440	445
30	Ser Ser Asp Asp Phe Cys	Arg Met Pro Pro Pro	Cys Ser Asp Ser Cys
	450	455	460
	Asp Phe Asp Asp Pro Arg	Leu Leu Lys Asn Ile	Glu Asp Arg His Pro
	465	470	475
	Thr Ala Pro Cys Ile Gln	Glu Phe Leu Thr	Leu Leu Ala Val Cys His
	485	490	495
35	Thr Val Val Pro Glu Lys	Asp Gly Asp Asn Ile	Ile Tyr Gln Ala Ser
	500	505	510
	Ser Pro Asp Glu Ala Ala	Leu Val Lys Gly	Ala Lys Lys Leu Gly Phe
	515	520	525
	Val Phe Thr Ala Arg Thr	Pro Phe Ser Val Ile	Ile Glu Ala Met Gly
40	530	535	540
	Gln Glu Gln Thr Phe Gly	Ile Leu Asn Val Leu	Glu Phe Ser Ser Asp
	545	550	555
	Arg Lys Arg Met Ser Val	Ile Val Arg Thr	Pro Ser Gly Arg Leu Arg
	565	570	575
45	Leu Tyr Cys Lys Gly	Ala Asp Asn Val Ile	Phe Glu Arg Leu Ser Lys
	580	585	590
	Asp Ser Lys Tyr Met Glu	Glu Thr Leu Cys His	Leu Glu Tyr Phe Ala
	595	600	605
	Thr Glu Gly Leu Arg	Thr Leu Cys Val Ala	Tyr Ala Asp Leu Ser Glu
	610	615	620
50	Asn Glu Tyr Glu Glu	Trp Leu Lys Val Tyr	Gln Glu Ala Ser Thr Ile
	625	630	635
	Leu Lys Asp Arg Ala	Gln Arg Leu Glu Glu	Cys Tyr Glu Ile Ile Glu
	645	650	655
	Lys Asn Leu Leu Leu	Gly Ala Thr Ala Ile	Glu Asp Arg Leu Gln
	660	665	670
55	Ala Gly Val Pro Glu Thr	Ile Ala Thr Leu Leu	Lys Ala Glu Ile Lys
	675	680	685

## EP 1 225 182 A2

Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr Ala Ile Asn Ile Gly  
 690 695 700  
 Tyr Ser Cys Arg Leu Val Ser Gln Asn Met Ala Leu Ile Leu Leu Lys  
 705 710 715 720  
 5 Glu Asp Ser Leu Asp Ala Thr Arg Ala Ala Ile Thr Gln His Cys Thr  
 725 730 735  
 Asp Leu Gly Asn Leu Leu Gly Lys Glu Asn Asp Val Ala Leu Ile Ile  
 740 745 750  
 10 Asp Gly His Thr Leu Lys Tyr Ala Leu Ser Phe Glu Val Arg Arg Ser  
 755 760 765  
 Phe Leu Asp Leu Ala Leu Ser Cys Lys Ala Val Ile Cys Cys Arg Val  
 770 775 780  
 Ser Pro Leu Gln Lys Ser Glu Ile Val Asp Val Val Lys Lys Arg Val  
 785 790 795 800  
 15 Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Gly Met  
 805 810 815  
 Ile Gln Thr Ala His Val Gly Val Gly Ile Ser Gly Asn Glu Gly Met  
 820 825 830  
 Gln Ala Thr Asn Asn Ser Asp Tyr Ala Ile Ala Gln Phe Ser Tyr Leu  
 835 840 845  
 20 Glu Lys Leu Leu Leu Val His Gly Ala Trp Ser Tyr Asn Arg Val Thr  
 850 855 860  
 Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Val Val Leu Tyr Ile Ile  
 865 870 875 880  
 Glu Leu Trp Phe Ala Phe Val Asn Gly Phe Ser Gly Gln Ile Leu Phe  
 885 890 895  
 25 Glu Arg Trp Cys Ile Gly Leu Tyr Asn Val Ile Phe Thr Ala Leu Pro  
 900 905 910  
 Pro Phe Thr Leu Gly Ile Phe Glu Arg Ser Cys Thr Gln Glu Ser Met  
 915 920 925  
 Leu Arg Phe Pro Gln Leu Tyr Lys Ile Thr Gln Asn Gly Glu Gly Phe  
 930 935 940  
 30 Asn Thr Lys Val Phe Trp Gly His Cys Ile Asn Ala Leu Val His Ser  
 945 950 955 960  
 Leu Ile Leu Phe Trp Phe Pro Met Lys Ala Leu Glu His Asp Thr Val  
 965 970 975  
 Leu Thr Ser Gly His Ala Thr Asp Tyr Leu Phe Val Gly Asn Ile Val  
 35 980 985 990  
 Tyr Thr Tyr Val Val Val Thr Val Cys Leu Lys Ala Gly Leu Glu Thr  
 995 1000 1005  
 Thr Ala Trp Thr Lys Phe Ser His Leu Ala Val Trp Gly Ser Met Leu  
 1010 1015 1020  
 40 Thr Trp Leu Val Phe Phe Gly Ile Tyr Ser Thr Ile Trp Pro Thr Ile  
 1025 1030 1035 1040  
 Pro Ile Ala Pro Asp Met Arg Gly Gln Ala Thr Met Val Leu Ser Ser  
 1045 1050 1055  
 Ala His Phe Trp Leu Gly Leu Phe Leu Val Pro Thr Ala Cys Leu Ile  
 1060 1065 1070  
 45 Glu Asp Val Ala Trp Arg Ala Ala Lys His Thr Cys Lys Lys Thr Leu  
 1075 1080 1085  
 Leu Glu Glu Val Gln Glu Leu Glu Thr Lys Ser Arg Val Leu Gly Lys  
 1090 1095 1100  
 50 Ala Val Leu Arg Asp Ser Asn Gly Lys Arg Leu Asn Glu Arg Asp Arg  
 1105 1110 1115 1120  
 Leu Ile Lys Arg Leu Gly Arg Lys Thr Pro Pro Thr Leu Phe Arg Gly  
 1125 1130 1135  
 Ser Ser Leu Gln Gln Gly Val Pro His Gly Tyr Ala Phe Ser Gln Glu  
 1140 1145 1150  
 55 Glu His Gly Ala Val Ser Gln Glu Val Ile Arg Ala Tyr Asp Thr  
 1155 1160 1165  
 Thr Lys Lys Ser Arg Lys Lys

EP 1 225 182 A2

1170

1175

5 <210> 3  
 <211> 3531  
 <212> DNA  
 <213> Homo sapiens  
 10 <220>  
 <221> CDS  
 <222> (1)...(3531)  
 15 <400> 3  
 atg tcc ctg ccg cgg agg tcg agg atc cgc tcg tcc gtg gga cct gtt 48  
 Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser Ser Val Gly Pro Val  
 1 5 10 15  
 20 cgt tct tct ttg ggc tat aag aag gca gat gat gag atg tcc cgg gcc 96  
 Arg Ser Ser Leu Gly Tyr Lys Lys Ala Asp Asp Glu Met Ser Arg Ala  
 25 20 25 30  
 25 acg tct gtt gga gac cag ctg gag gca ccc gcc cgc acc att tac ctc 144  
 Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala Arg Thr Ile Tyr Leu  
 30 35 40 45  
 30 aac caa ccg cat ctc aac aaa ttc cgc gac aac cag atc agt acg gcc 192  
 Asn Gln Pro His Leu Asn Lys Phe Arg Asp Asn Gln Ile Ser Thr Ala  
 35 50 55 60  
 35 aag tac agc gtg ttg aca ttt cta cct cga ttc ttg tat gag cag att 240  
 Lys Tyr Ser Val Leu Thr Phe Leu Pro Arg Phe Leu Tyr Glu Gln Ile  
 40 65 70 75 80  
 40 aga aga gct gct aat gcc ttc ttt ctc ttc att gcc tta tta cag caa 288  
 Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile Ala Leu Leu Gln Gln  
 45 85 90 95  
 45 att cca gat gta tct cca aca gga aga tat acc acc ctg gtg cca ttg 336  
 Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr Leu Val Pro Leu  
 50 100 105 110  
 50 atc att att cta aca att gca ggc atc aag gag att gta gaa gat ttt 384  
 Ile Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu Ile Val Glu Asp Phe  
 55 115 120 125  
 55 aag cga cac aag gca gac aat gca gtt aac aaa aag aaa aca ata gtg 432  
 Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys Lys Lys Thr Ile Val  
 60 130 135 140  
 60 tta aga aat ggt atg tgg cat acc att atg tgg aaa gag gtg gca gtg 480  
 Leu Arg Asn Gly Met Trp His Thr Ile Met Trp Lys Glu Val Ala Val  
 65 145 150 155 160  
 65 gga gac att gtg aag gtc gtc aat ggg cag tat ctt cca gca gat gtg 528  
 Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr Leu Pro Ala Asp Val  
 70 165 170 175  
 70 gtc ctg ctg tca tcc agt gaa cct cag gca atg tgt tat gtt gaa aca 576  
 Val Leu Leu Ser Ser Glu Pro Gln Ala Met Cys Tyr Val Glu Thr  
 75 180 185 190

EP 1 225 182 A2

5	gct aat ctg gat ggg gag acg aac ctt aaa ata cgt cag ggt ttg agt Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile Arg Gln Gly Leu Ser 195 200 205	624
10	cac act gct gac atg caa aca cgt gaa gtt ctg atg aag tta tct gga His Thr Ala Asp Met Gln Thr Arg Glu Val Leu Met Lys Leu Ser Gly 210 215 220	672
15	act ata gag tgc gaa ggg ccc aac cgc cac ctc tat gac ttc act gga Thr Ile Glu Cys Glu Gly Pro Asn Arg His Leu Tyr Asp Phe Thr Gly 225 230 235 240	720
20	aac ttg aac tta gat ggg aaa agc ctt gtt gcc ctt ggg cct gac cag Asn Leu Asn Leu Asp Gly Lys Ser Leu Val Ala Leu Gly Pro Asp Gln 245 250 255	768
25	atc tta tta aga ggt aca cag ctt aga aat act cag tgg gtc ttt ggc Ile Leu Leu Arg Gly Thr Gln Leu Arg Asn Thr Gln Trp Val Phe Gly 260 265 270	816
30	ata gtt gtt tat act gga cac gac acc aaa ctc atg cag aat tca acc Ile Val Val Tyr Thr Gly His Asp Thr Lys Leu Met Gln Asn Ser Thr 275 280 285	864
35	aaa gcg cct ctc aag aga tca aat gtt gag aag gtg act aac gtg cag Lys Ala Pro Leu Lys Arg Ser Asn Val Glu Lys Val Thr Asn Val Gln 290 295 300	912
40	atc ctg gtg ttg ttt ggc aac ctc ttg gtc atg gcc ttg gtg agc tcg Ile Leu Val Leu Phe Gly Ile Leu Leu Val Met Ala Leu Val Ser Ser 305 310 315 320	960
45	gcg ggg gcc ctg tac tgg aac agg tct cat ggt gaa aag aac tgg tac Ala Gly Ala Leu Tyr Trp Asn Arg Ser His Gly Glu Lys Asn Trp Tyr 325 330 335	1008
50	atc aag aag atg gac acc acc tca gat aat ttt gga tac aac cta ctg Ile Lys Lys Met Asp Thr Thr Ser Asp Asn Phe Gly Tyr Asn Leu Leu 340 345 350	1056
55	acg ttc atc atc tta tac aac aat ctt att ccc atc agt ctg ttg gtg Thr Phe Ile Ile Leu Tyr Asn Asn Leu Ile Pro Ile Ser Leu Leu Val 355 360 365	1104
60	act ctt gag gtt gtg aag tat act caa gcc ctt ttc ata aac tgg gac Thr Leu Glu Val Val Lys Tyr Thr Gln Ala Leu Phe Ile Asn Trp Asp 370 375 380	1152
65	aca gat atg tat tat ata gga aat gac act cct gcc atg gcc agg aca Thr Asp Met Tyr Tyr Ile Gly Asn Asp Thr Pro Ala Met Ala Arg Thr 385 390 395 400	1200
70	tca aac ctt aat gaa gag ctt ggg cag gtg aaa tat ctc ttt tct gac Ser Asn Leu Asn Glu Leu Gly Gln Val Lys Tyr Leu Phe Ser Asp 405 410 415	1248
75	aag act gga acg ctt aca tgc aat atc atg aac ttt aag aag tgc agc Lys Thr Gly Thr Leu Thr Cys Asn Ile Met Asn Phe Lys Lys Cys Ser 420 425 430	1296
80	att gcc gga gta acc tat ggt cac ttc cca gaa ttg gca aga gag ccg	1344

EP 1 225 182 A2

	Ile Ala Gly Val Thr Tyr Gly His Phe Pro Glu Leu Ala Arg Glu Pro	
	435 440 445	
5	tct tca gat gac ttc tgt cgg atg cct cct ccc tgt agt gat tcc tgt Ser Ser Asp Asp Phe Cys Arg Met Pro Pro Pro Cys Ser Asp Ser Cys	1392
	450 455 460	
10	gac ttt gat gac ccc agg ctg ttg aag aac att gag gat cgc cat ccc Asp Phe Asp Asp Pro Arg Leu Leu Lys Asn Ile Glu Asp Arg His Pro	1440
	465 470 475 480	
	aca gcc cct tgc att cag gag ttc ctc acc ctt ctg gcc gtg tgc cac Thr Ala Pro Cys Ile Gln Glu Phe Leu Thr Leu Leu Ala Val Cys His	1488
	485 490 495	
15	acg gtt gtt cct gag aag gat gga gat aac atc atc tac cag gcc tct Thr Val Val Pro Glu Lys Asp Gly Asp Asn Ile Ile Tyr Gln Ala Ser	1536
	500 505 510	
20	tcc cca gat gaa gct gct ttg gtg aaa gga gct aaa aag ctg ggc ttt Ser Pro Asp Glu Ala Ala Leu Val Lys Gly Ala Lys Lys Leu Gly Phe	1584
	515 520 525	
25	gtc ttc aca gcc aga aca cca ttc tca gtc atc ata gaa gcg atg gga Val Phe Thr Ala Arg Thr Pro Phe Ser Val Ile Ile Glu Ala Met Gly	1632
	530 535 540	
	cag gaa caa aca ttc gga atc ctt aat gtc ctg gaa ttt tct agt gac Gln Glu Gln Thr Phe Gly Ile Leu Asn Val Leu Glu Phe Ser Ser Asp	1680
	545 550 555 560	
30	aga aaa aga atg tct gta att gtt cga act cct tca gga cga ctt cgg Arg Lys Arg Met Ser Val Ile Val Arg Thr Pro Ser Gly Arg Leu Arg	1728
	565 570 575	
35	ctt tac tgt aaa ggg gct gat aat gtg att ttt gag aga ctt tca aaa Leu Tyr Cys Lys Gly Ala Asp Asn Val Ile Phe Glu Arg Leu Ser Lys	1776
	580 585 590	
	gac tca aaa tat atg gag gaa aca tta tgc cat ctg gaa tac ttt gcc Asp Ser Lys Tyr Met Glu Glu Thr Leu Cys His Leu Glu Tyr Phe Ala	1824
	595 600 605	
40	acg gaa ggc ttg cgg act ctc tgt gtg gct tat gct gat ctc tct gag Thr Glu Gly Leu Arg Thr Leu Cys Val Ala Tyr Ala Asp Leu Ser Glu	1872
	610 615 620	
45	aat gag tat gag gag tgg ctg aaa gtc tat cag gaa gcc agc acc ata Asn Glu Tyr Glu Glu Trp Leu Lys Val Tyr Gln Glu Ala Ser Thr Ile	1920
	625 630 635 640	
	ttg aag gac aga gct caa cgg ttg gaa gag tgt tac gag atc att gag Leu Lys Asp Arg Ala Gln Arg Leu Glu Glu Cys Tyr Glu Ile Ile Glu	1968
	645 650 655	
50	aag aat ttg ctg cta ctt gga gcc aca gcc ata gaa gat cgc ctt caa Lys Asn Leu Leu Leu Gly Ala Thr Ala Ile Glu Asp Arg Leu Gln	2016
	660 665 670	
55	gca gga gtt cca gaa acc atc gca aca ctg ttg aag gca gaa att aaa Ala Gly Val Pro Glu Thr Ile Ala Thr Leu Leu Lys Ala Glu Ile Lys	2064

EP 1 225 182 A2

	675	680	685	
5	ata tgg gtg ttg aca gga gac aaa caa gaa act gcg att aat ata ggg Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr Ala Ile Asn Ile Gly 690	695	700	2112
10	tat tcc tgc cga ttg gta tcg cag aat atg gcc ctt atc cta ttg aag Tyr Ser Cys Arg Leu Val Ser Gln Asn Met Ala Leu Ile Leu Lys 705	710	715	2160
15	gag gac tct ttg gat gcc aca agg gca gcc att act cag cac tgc act Glu Asp Ser Leu Asp Ala Thr Arg Ala Ala Ile Thr Gln His Cys Thr 725	730	735	2208
20	gac ctt ggg aat ttg ctg ggc aag gaa aat gac gtg gcc ctc atc atc Asp Leu Gly Asn Leu Leu Gly Lys Glu Asn Asp Val Ala Leu Ile Ile 740	745	750	2256
25	gat ggc cac acc ctg aag tac gcg ctc tcc ttc gaa gtc cgg agg agt Asp Gly His Thr Leu Lys Tyr Ala Leu Ser Phe Glu Val Arg Arg Ser 755	760	765	2304
30	ttc ctg gat ttg gca ctc tcg tgc aaa gcg gtc ata tgc tgc aga gtg Phe Leu Asp Leu Ala Leu Ser Cys Lys Ala Val Ile Cys Cys Arg Val 770	775	780	2352
35	tct cct ctg cag aag tct gag ata gtg gat gtg gtg aag aag cgg gtg Ser Pro Leu Gln Lys Ser Glu Ile Val Asp Val Val Lys Lys Arg Val 785	790	795	2400
40	aag gcc atc acc ctc gcc atc gga gac ggc gcc aac gat gtc ggg atg Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Gly Met 805	810	815	2448
45	atc cag aca gcc cac gtg ggt gtg gga atc agt ggg aat gaa ggc atg Ile Gln Thr Ala His Val Gly Val Gly Ile Ser Gly Asn Glu Gly Met 820	825	830	2496
50	cag gcc acc aac aac tcg gat tac gca cag ttt tcc tac tta Gln Ala Thr Asn Asn Ser Asp Tyr Ala Ile Ala Gln Phe Ser Tyr Leu 835	840	845	2544
55	gag aag ctt ctg ttg gtt cat gga gcc tgg agc tac aac cgg gtg acc Glu Lys Leu Leu Val His Gly Ala Trp Ser Tyr Asn Arg Val Thr 850	855	860	2592
60	aag tgc atc ttg tac tgc ttc tat aag aac gtg gtc ctg tat att att Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Val Val Leu Tyr Ile Ile 865	870	875	2640
65	gag ctt tgg ttc gcc ttt gtt aat gga ttt tct ggg cag att tta ttt Glu Leu Trp Phe Ala Phe Val Asn Gly Phe Ser Gly Gln Ile Leu Phe 885	890	895	2688
70	gaa cgt tgg tgc atc ggc ctg tac aat gtg att ttc acc gct ttg ccg Glu Arg Trp Cys Ile Gly Leu Tyr Asn Val Ile Phe Thr Ala Leu Pro 900	905	910	2736
75	ccc ttc act ctg gga atc ttt gag agg tct tgc act cag gag agc atg Pro Phe Thr Leu Gly Ile Phe Glu Arg Ser Cys Thr Gln Glu Ser Met 915	920	925	2784

EP 1 225 182 A2

5	ctc agg ttt ccc cag ctc tac aaa atc acc cag aat ggc gaa ggc ttc Leu Arg Phe Pro Gln Leu Tyr Lys Ile Thr Gln Asn Gly Glu Gly Phe 930 935 940	2832
10	aac aca aag gtt ttc tgg ggt cac tgc atc aac gcc ttg gtc cac tcc Asn Thr Lys Val Phe Trp Gly His Cys Ile Asn Ala Leu Val His Ser 945 950 955 960	2880
15	ctc atc ctc ttc tgg ttt ccc atg aaa gct ctg gag cat gat act gtg Leu Ile Leu Phe Trp Phe Pro Met Lys Ala Leu Glu His Asp Thr Val 965 970 975	2928
20	ttg aca agt ggt cat gct acc gac tat tta ttt gtt gga aat att gtt Leu Thr Ser Gly His Ala Thr Asp Tyr Leu Phe Val Gly Asn Ile Val 980 985 990	2976
25	tac aca tat gtt gtt act gtt tgt ctg aaa gct ggt ttg gag acc Tyr Thr Tyr Val Val Val Thr Val Cys Leu Lys Ala Gly Leu Glu Thr 995 1000 1005	3024
30	aca gct tgg act aaa ttc agt cat ctg gct gtc tgg gga agc atg ctg Thr Ala Trp Thr Lys Phe Ser His Leu Ala Val Trp Gly Ser Met Leu 1010 1015 1020	3072
35	acc tgg ctg gtg ttt ttt ggc atc tac tcg acc atc tgg ccc acc att Thr Trp Leu Val Phe Phe Gly Ile Tyr Ser Thr Ile Trp Pro Thr Ile 1025 1030 1035 1040	3120
40	ccc att gct cca gat atg aga gga cag gca act atg gtc ctg agc tcc Pro Ile Ala Pro Asp Met Arg Gly Gln Ala Thr Met Val Leu Ser Ser 1045 1050 1055	3168
45	gca cac ttc tgg ttg gga tta ttt ctg gtt cct act gcc tgt ttg att Ala His Phe Trp Leu Gly Leu Phe Leu Val Pro Thr Ala Cys Leu Ile 1060 1065 1070	3216
50	gaa gat gtg gca tgg aga gca gcc aag cac acc tgc aaa aag aca ttg Glu Asp Val Ala Trp Arg Ala Ala Lys His Thr Cys Lys Lys Thr Leu 1075 1080 1085	3264
55	ctg gag gag gtg cag gag ctg gaa acc aag tct cga gtc ctg gga aaa Leu Glu Glu Val Gln Glu Leu Glu Thr Lys Ser Arg Val Leu Gly Lys 1090 1095 1100	3312
60	gcg gtg ctg cgg gat agc aat gga aag agg ctg aac gag cgc gac cgc Ala Val Leu Arg Asp Ser Asn Gly Lys Arg Leu Asn Glu Arg Asp Arg 1105 1110 1115 1120	3360
65	ctg atc aag agg ctg ggc cgg aag acg ccc ccg acg ctg ttc cgg ggc Leu Ile Lys Arg Leu Gly Arg Lys Thr Pro Pro Thr Leu Phe Arg Gly 1125 1130 1135	3408
70	agc tcc ctg cag cag ggc gtc ccg cat ggg tat gct ttt tct caa gaa Ser Ser Leu Gln Gln Gly Val Pro His Gly Tyr Ala Phe Ser Gln Glu 1140 1145 1150	3456
75	gaa cac gga gct gtt agt cag gaa gaa gtc atc cgt gct tat gac acc Glu His Gly Ala Val Ser Gln Glu Glu Val Ile Arg Ala Tyr Asp Thr 1155 1160 1165	3504

EP 1 225 182 A2

acc aaa aag aaa tcc agg aag aaa taa	3531
Thr Lys Lys Lys Ser Arg Lys Lys *	
1170	1175
5	
<210> 4	
<211> 2997	
<212> DNA	
<213> homo sapiens	
10	
<220>	
<221> CDS	
<222> (20)...(2671)	
15	
<400> 4	
cctggacaaa gctcttaag atg tcc ctg ccg cgg agg tcg agg atc cgc tcg	52
Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser	
1	5
10	10
tcc gtg gga cct gtt cgt tct ttg ggc tat aag aag gca gat gat	100
Ser Val Gly Pro Val Arg Ser Ser Leu Gly Tyr Lys Lys Ala Asp Asp	
15	20
20	25
gag atg tcc cgg gcc acg tct gtt gga gac cag ctg gag gca ccc gcc	148
Glu Met Ser Arg Ala Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala	
30	35
25	40
cgc acc att tac ctc aac caa ccg cat ctc aac aaa ttc cgc gac aac	196
Arg Thr Ile Tyr Leu Asn Gln Pro His Leu Asn Lys Phe Arg Asp Asn	
45	50
55	55
cag atc agt acg gcc aag tac agc gtg ttg aca ttt cta cct cga ttc	244
Gln Ile Ser Thr Ala Lys Tyr Ser Val Leu Thr Phe Leu Pro Arg Phe	
60	65
70	75
ttg tat gag cag att aga aga gct gct aat gcc ttc ttt ctc ttc att	292
Leu Tyr Glu Gln Ile Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile	
80	85
90	90
gcc tta tta cag caa att cca gat gta tct cca aca gga aga tat acc	340
Ala Leu Leu Gln Gln Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr	
95	100
105	105
40	
acc ctg gtg cca ttg atc att att tta aca att gca ggc atc aag gag	388
Thr Leu Val Pro Leu Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu	
110	115
120	120
45	
att gta gaa gat ttt aag cga cac aag gca gac aat gca gtt aac aaa	436
Ile Val Glu Asp Phe Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys	
125	130
135	135
50	
aag aaa aca ata gtg tta aga aat ggt atg tgg cat acc att atg tgg	484
Lys Lys Thr Ile Val Leu Arg Asn Gly Met Trp His Thr Ile Met Trp	
140	145
150	155
55	
aaa gag gtg gca gtg gga gac att gtg aag gtc gtc aat ggg cag tat	532
Lys Glu Val Ala Val Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr	
160	165
170	170
ctt cca gca gat gtg gtc ctg ctg tca tcc agt gaa cct cag gca atg	580
Leu Pro Ala Asp Val Val Leu Leu Ser Ser Glu Pro Gln Ala Met	

## EP 1 225 182 A2

	175	180	185	
5	tgt tat gtt gaa aca gct aat ctg gat ggg gag acg aac ctt aaa ata Cys Tyr Val Glu Thr Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile 190 195 200			628
	cgt cag ggt ttg agt cac act gct gac atg caa aca cgt gaa gtt ctg Arg Gln Gly Leu Ser His Thr Ala Asp Met Gln Thr Arg Glu Val Leu 205 210 215			676
10	atg aag tta tct gga act ata gag cgt gaa ggg ccc aac cgc cac ctc Met Lys Leu Ser Gly Thr Ile Glu Arg Glu Gly Pro Asn Arg His Leu 220 225 230 235			724
15	tat gac ttc act gga aac ttg aac tta gat gga aaa gcc ttg ccc Tyr Asp Phe Thr Gly Asn Leu Asn Leu Asp Gly Lys Ala Leu Pro 240 245 250			772
20	ttg ggc gct gac cag atc tta tta aga ggt aca cag ctt aga aat act Leu Gly Ala Asp Gln Ile Leu Leu Arg Gly Thr Gln Leu Arg Asn Thr 255 260 265			820
	cag tgg gtc ttt ggc ata gtt gtt tat act gga cac gac acc aaa ctc Gln Trp Val Phe Gly Ile Val Val Tyr Thr Gly His Asp Thr Lys Leu 270 275 280			868
25	atg cag aat tca acc aaa gca cct ctc aag aga tca aat gtt gag aag Met Gln Asn Ser Thr Lys Ala Pro Leu Lys Arg Ser Asn Val Glu Lys 285 290 295			916
30	gtg act aac gtg cag atc ctg gtg ttt ggc atc ctc ttg gtc atg Val Thr Asn Val Gln Ile Leu Val Leu Phe Gly Ile Leu Leu Val Met 300 305 310 315			964
	gcc ttg gtg agc tcg gcg ggg gcc ctg tac tgg aac agg tct cat ggt Ala Leu Val Ser Ser Ala Gly Ala Leu Tyr Trp Asn Arg Ser His Gly 320 325 330			1012
35	gaa aag aac tgg tac atc aag aag atg gac acc acc tca gat aat ttt Glu Lys Asn Trp Tyr Ile Lys Lys Met Asp Thr Thr Ser Asp Asn Phe 335 340 345			1060
40	gga tac aac cta ctg acg ttc atc atc tta tac aac aat ctt att ccc Gly Tyr Asn Leu Leu Thr Phe Ile Ile Leu Tyr Asn Asn Leu Ile Pro 350 355 360			1108
	atc agt ctg ttg gtg act ctt gag gtt gtg aag tat act caa gcc ctt Ile Ser Leu Leu Val Thr Leu Glu Val Val Lys Tyr Thr Gln Ala Leu 365 370 375			1156
45	ttc ata aac tgg gac aca gat atg tat tat ata gga aat gac act cct Phe Ile Asn Trp Asp Thr Asp Met Tyr Tyr Ile Gly Asn Asp Thr Pro 380 385 390 395			1204
50	gcc atg gcc agg aca tca aac ctt aat gaa gag ctt ggg cag gtg aaa Ala Met Ala Arg Thr Ser Asn Leu Asn Glu Glu Leu Gly Gln Val Lys 400 405 410			1252
55	tat ctg ttt tct gac aag act gga acg ctt aca tgc aat atc atg aac Tyr Leu Phe Ser Asp Lys Thr Gly Thr Leu Thr Cys Asn Ile Met Asn 415 420 425			1300

EP 1 225 182 A2

	ttt aag aag tgc agc att gcc 'gga gta acc tat ggt cac ttc cca gaa Phe Lys Lys Cys Ser Ile Ala Gly Val Thr Tyr Gly His Phe Pro Glu 430 435 440	1348
5	ttg gca aga gag ccg tct tca gat gac ttc tgt cgg atg cct cct ccc Leu Ala Arg Glu Pro Ser Ser Asp Asp Phe Cys Arg Met Pro Pro Pro 445 450 455	1396
10	tgt agt gat tcc tgt gac ttt gat gac ccc agg ctg ttg aag aac att Cys Ser Asp Ser Cys Asp Phe Asp Asp Pro Arg Leu Leu Lys Asn Ile 460 465 470 475	1444
15	gag gat cgc cat ccc aca gcc cct tgc att cag gag ttc ctc acc ctt Glu Asp Arg His Pro Thr Ala Pro Cys Ile Gln Glu Phe Leu Thr Leu 480 485 490	1492
20	ctg gcc gtg tgc cac acg gtt gtt cct gag aag gat gga gat aac atc Leu Ala Val Cys His Thr Val Val Pro Glu Lys Asp Gly Asp Asn Ile 495 500 505	1540
25	atc tac cag gcc tct tcc cca gat gaa gct gct ttg gtg aaa gga gct Ile Tyr Gln Ala Ser Ser Pro Asp Glu Ala Ala Leu Val Lys Gly Ala 510 515 520	1588
30	aaa aag ctg ggc ttt gtc ttc aca gcc aga aca cca ttc tca gtc atc Lys Lys Leu Gly Phe Val Phe Thr Ala Arg Thr Pro Phe Ser Val Ile 525 530 535	1636
35	ata gaa gcg atg gga cag gaa caa aca ttc gga atc ctt aat gtc ctg Ile Glu Ala Met Gly Gln Glu Gln Thr Phe Gly Ile Leu Asn Val Leu 540 545 550 555	1684
40	gaa ttt tct agt gac aga aaa aga atg tct gta att gtt cga act cct Glu Phe Ser Ser Asp Arg Lys Arg Met Ser Val Ile Val Arg Thr Pro 560 565 570	1732
45	tca gga cga ctt cgg ctt tac tgt aaa ggg gct gat aat gtg att ttt Ser Gly Arg Leu Arg Leu Tyr Cys Lys Gly Ala Asp Asn Val Ile Phe 575 580 585	1780
50	gag aga ctt tca aaa gac tca aaa tat atg gag gaa aca tta tgc cat Glu Arg Leu Ser Lys Asp Ser Lys Tyr Met Glu Glu Thr Leu Cys His 590 595 600	1828
55	ctg gaa tac ttt gcc acg gaa ggc ttg cgg act ctc tgt gtg gct tat Leu Glu Tyr Phe Ala Thr Glu Gly Leu Arg Thr Leu Cys Val Ala Tyr 605 610 615	1876
	gct gat ctc tct gag aat gag tat gag gag tgg ctg aaa gtc tat cag Ala Asp Leu Ser Glu Asn Glu Tyr Glu Glu Trp Leu Lys Val Tyr Gln 620 625 630 635	1924
	gaa gcc agc acc ata ttg aag gac aga gct caa cgg ttg gaa gag tgt Glu Ala Ser Thr Ile Leu Lys Asp Arg Ala Gln Arg Leu Glu Glu Cys 640 645 650	1972
	tac gag atc att gag aag aat ttg ctg cta ctt gga gcc aca gcc ata Tyr Glu Ile Ile Glu Lys Asn Leu Leu Leu Leu Gly Ala Thr Ala Ile 655 660 665	2020

EP 1 225 182 A2

	gaa gat cgc ctt caa gca gga gtt cca gaa acc atc gca aca ctg ttg Glu Asp Arg Leu Gln Ala Gly Val Pro Glu Thr Ile Ala Thr Leu Leu 670 675 680	2068
5	aag gca gaa att aaa ata tgg gtg ttg aca gga gac aaa caa gaa act Lys Ala Glu Ile Lys Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr 685 690 695	2116
10	gcg att aat ata ggg tat tcc tgc cga ttg gta tcg cag aat atg gcc Ala Ile Asn Ile Gly Tyr Ser Cys Arg Leu Val Ser Gln Asn Met Ala 700 705 710 715	2164
15	ctt atc cta ttg aag gag gac tct ttg gat gcc aca agg gca gcc att Leu Ile Leu Leu Lys Glu Asp Ser Leu Asp Ala Thr Arg Ala Ala Ile 720 725 730	2212
	act cag cac tgc act gac ctt ggg aat ttg ctg ggc aag gaa aat gac Thr Gln His Cys Thr Asp Leu Gly Asn Leu Leu Gly Lys Glu Asn Asp 735 740 745	2260
20	gtg gcc ctc atc atc gat ggc cac acc ctg aag tac gcg ctc tcc ttc Val Ala Leu Ile Ile Asp Gly His Thr Leu Lys Tyr Ala Leu Ser Phe 750 755 760	2308
25	gaa gtc cgg agg agt ttc ctg gat ttg gca ctc tcg tgc aaa gcg gtc Glu Val Arg Arg Ser Phe Leu Asp Leu Ala Leu Ser Cys Lys Ala Val 765 770 775	2356
	ata tgc tgc aga gtg tct cct ctg cag aag tct gag ata gtg gat gtg Ile Cys Cys Arg Val Ser Pro Leu Gln Lys Ser Glu Ile Val Asp Val 780 785 790 795	2404
30	gtg aag aag cgg gtg aag gcc atc acc ctc gcc atc gga gac ggc gcc Val Lys Lys Arg Val Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala 800 805 810	2452
35	aac gat gtc ggg atg atc cag aca gcc cac gtg ggt gtg gga atc agt Asn Asp Val Gly Met Ile Gln Thr Ala His Val Gly Val Gly Ile Ser 815 820 825	2500
	ggg aat gaa ggc atg cag gcc acc aac aac tcg gat tac gcc atc gca Gly Asn Glu Gly Met Gln Ala Thr Asn Asn Ser Asp Tyr Ala Ile Ala 830 835 840	2548
40	cag ttt tcc tac tta gag aag ctt ctg ttg gtt cat gga gcc tgg agc Gln Phe Ser Tyr Leu Glu Lys Leu Leu Val His Gly Ala Trp Ser 845 850 855	2596
45	tac aac cgg gtg acc aag tgc atc ttg tac tgc ttc tat aag aac gtg Tyr Asn Arg Val Thr Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Val 860 865 870 875	2644
	gtc ctg tat att att gag aaa att tag aaaaacgacct aatcacatg Val Leu Tyr Ile Ile Glu Lys Ile *	2691
50	tcagcagata gaagatttca aataactttg caatacagca ccatattatt ttggcatac ggttcagaag aaatctaaag tattgatggc tcttgcgtta atgaagcttt acattcttt taatgtgtga actattata gatttctcag gactgagatc tatagaaaatg agtaatagga atagaaaattt aaacaaaattt aagtaagatt tccatattacc atttctgtga agtataacaa gtgatcaata aaataaaca aactacatta aattatggta aaaaaaaaaaaa aaaaaaaggc	2751 2811 2871 2931 2991

ggccgc

2997

5 <210> 5  
 <211> 883  
 <212> PRT  
 <213> homo sapiens

10 <400> 5  
 Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser Ser Val Gly Pro Val  
 1 5 10 15  
 Arg Ser Ser Leu Gly Tyr Lys Ala Asp Asp Glu Met Ser Arg Ala  
 20 25 30  
 Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala Arg Thr Ile Tyr Leu  
 35 40 45  
 15 Asn Gln Pro His Leu Asn Lys Phe Arg Asp Asn Gln Ile Ser Thr Ala  
 50 55 60  
 Lys Tyr Ser Val Leu Thr Phe Lys Pro Arg Phe Leu Tyr Glu Gln Ile  
 65 70 75 80  
 Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile Ala Leu Leu Gln Gln  
 85 90 95  
 20 Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr Thr Leu Val Pro Leu  
 100 105 110  
 Ile Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu Ile Val Glu Asp Phe  
 115 120 125  
 Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys Lys Lys Thr Ile Val  
 130 135 140  
 25 Leu Arg Asn Gly Met Trp His Thr Ile Met Trp Lys Glu Val Ala Val  
 145 150 155 160  
 Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr Leu Pro Ala Asp Val  
 165 170 175  
 Val Leu Leu Ser Ser Glu Pro Gln Ala Met Cys Tyr Val Glu Thr  
 180 185 190  
 30 Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile Arg Gln Gly Leu Ser  
 195 200 205  
 His Thr Ala Asp Met Gln Thr Arg Glu Val Leu Met Lys Leu Ser Gly  
 210 215 220  
 Thr Ile Glu Arg Glu Gly Pro Asn Arg His Leu Tyr Asp Phe Thr Gly  
 225 230 235 240  
 35 Asn Leu Asn Leu Asp Gly Lys Ala Leu Leu Pro Leu Gly Ala Asp Gln  
 245 250 255  
 Ile Leu Leu Arg Gly Thr Gln Leu Arg Asn Thr Gln Trp Val Phe Gly  
 260 265 270  
 40 Ile Val Val Tyr Thr Gly His Asp Thr Lys Leu Met Gln Asn Ser Thr  
 275 280 285  
 Lys Ala Pro Leu Lys Arg Ser Asn Val Glu Lys Val Thr Asn Val Gln  
 290 295 300  
 Ile Leu Val Leu Phe Gly Ile Leu Leu Val Met Ala Leu Val Ser Ser  
 305 310 315 320  
 45 Ala Gly Ala Leu Tyr Trp Asn Arg Ser His Gly Glu Lys Asn Trp Tyr  
 325 330 335  
 Ile Lys Lys Met Asp Thr Thr Ser Asp Asn Phe Gly Tyr Asn Leu Leu  
 340 345 350  
 Thr Phe Ile Leu Tyr Asn Asn Leu Ile Pro Ile Ser Leu Leu Val  
 355 360 365  
 50 Thr Leu Glu Val Val Lys Tyr Thr Gln Ala Leu Phe Ile Asn Trp Asp  
 370 375 380  
 Thr Asp Met Tyr Tyr Ile Gly Asn Asp Thr Pro Ala Met Ala Arg Thr  
 385 390 395 400  
 Ser Asn Leu Asn Glu Glu Leu Gly Gln Val Lys Tyr Leu Phe Ser Asp  
 405 410 415  
 55 Lys Thr Gly Thr Leu Thr Cys Asn Ile Met Asn Phe Lys Lys Cys Ser

## EP 1 225 182 A2

	420	425	430
	Ile Ala Gly Val Thr Tyr Gly His Phe Pro Glu Leu Ala Arg Glu Pro		
	435	440	445
5	Ser Ser Asp Asp Phe Cys Arg Met Pro Pro Pro Cys Ser Asp Ser Cys		
	450	455	460
	Asp Phe Asp Asp Pro Arg Leu Leu Lys Asn Ile Glu Asp Arg His Pro		
	465	470	475
10	480	485	490
	Thr Ala Pro Cys Ile Gln Glu Phe Leu Thr Leu Leu Ala Val Cys His		
	495	500	505
	510	515	520
15	Ser Pro Asp Glu Ala Ala Leu Val Lys Gly Ala Lys Lys Leu Gly Phe		
	525	530	535
	540	545	550
	555	560	560
20	Arg Lys Arg Met Ser Val Ile Val Arg Thr Pro Ser Gly Arg Leu Arg		
	565	570	575
	580	585	590
	Leu Tyr Cys Lys Gly Ala Asp Asn Val Ile Phe Glu Arg Leu Ser Lys		
25	595	600	605
	Asp Ser Lys Tyr Met Glu Glu Thr Leu Cys His Leu Glu Tyr Phe Ala		
	610	615	620
	625	630	635
30	640	645	650
	655	660	665
	670	675	680
	685	690	695
35	700	705	710
	720	725	730
	735	740	745
40	750	755	760
	765	770	775
	780	785	790
45	800	805	810
	815	820	825
	830	835	840
50	845	850	855
	860	865	870
55	880	Glu Lys Ile	

EP 1 225 182 A2

5 <210> 6  
 <211> 2652  
 <212> DNA  
 <213> homo sapiens  
  
 10 <220>  
 <221> CDS  
 <222> (1)...(2652)  
  
 15 <400> 6  
 atg tcc ctg ccg cgg agg tcg agg atc cgc tcg tcc gtg gga cct gtt 48  
 Met Ser Leu Pro Arg Arg Ser Arg Ile Arg Ser Ser Val Gly Pro Val  
 1 5 10 15  
  
 20 cgt tct tct ttg ggc tat aag aag gca gat gat gag atg tcc cgg gcc 96  
 Arg Ser Ser Leu Gly Tyr Lys Lys Ala Asp Asp Glu Met Ser Arg Ala  
 20 25 30  
  
 25 acg tct gtt gga gac cag ctg gag gca ccc gcc cgc acc att tac ctc 144  
 Thr Ser Val Gly Asp Gln Leu Glu Ala Pro Ala Arg Thr Ile Tyr Leu  
 20 35 40 45  
  
 30 aac caa ccg cat ctc aac aaa ttc cgc gac aac cag atc agt acg gcc 192  
 Asn Gln Pro His Leu Asn Lys Phe Arg Asp Asn Gln Ile Ser Thr Ala  
 50 55 60  
  
 35 aag tac agc gtg ttg aca ttt cta cct cga ttc ttg tat gag cag att 240  
 Lys Tyr Ser Val Leu Thr Phe Leu Pro Arg Phe Leu Tyr Glu Gln Ile  
 65 70 75 80  
  
 40 aga aga gct aat gcc ttc ttt ctc ttc att gcc tta tta cag caa 288  
 Arg Arg Ala Ala Asn Ala Phe Phe Leu Phe Ile Ala Leu Leu Gln Gln  
 85 90 95  
  
 45 att cca gat gta tct cca aca gga aga tat acc acc ctg gtg cca ttg 336  
 Ile Pro Asp Val Ser Pro Thr Gly Arg Tyr Thr Leu Val Pro Leu  
 100 105 110  
  
 50 atc att att tta aca att gca ggc atc aag gag att gta gaa gat ttt 384  
 Ile Ile Ile Leu Thr Ile Ala Gly Ile Lys Glu Ile Val Glu Asp Phe  
 115 120 125  
  
 55 aag cga cac aag gca gac aat gca gtt aac aaa aag aaa aca ata gtg 432  
 Lys Arg His Lys Ala Asp Asn Ala Val Asn Lys Lys Thr Ile Val  
 130 135 140  
  
 60 tta aga aat ggt atg tgg cat acc att atg tgg aaa gag gtg gca gtg 480  
 Leu Arg Asn Gly Met Trp His Thr Ile Met Trp Lys Glu Val Ala Val  
 145 150 155 160  
  
 65 gga gac att gtg aag gtc gtc aat ggg cag tat ctt cca gca gat gtg 528  
 Gly Asp Ile Val Lys Val Val Asn Gly Gln Tyr Leu Pro Ala Asp Val  
 165 170 175  
  
 70 gtc ctg ctg tca tcc agt gaa cct cag gca atg tgt tat gtt gaa aca 576  
 Val Leu Leu Ser Ser Ser Glu Pro Gln Ala Met Cys Tyr Val Glu Thr  
 180 185 190  
  
 75 gct aat ctg gat ggg gag acg aac ctt aaa ata cgt cag ggt ttg agt 624  
 Ala Asn Leu Asp Gly Glu Thr Asn Leu Lys Ile Arg Gln Gly Leu Ser  
 195 200 205

EP 1 225 182 A2

	cac act gct gac atg caa aca cgt gaa gtt ctg atg aag tta tct gga	672
	His Thr Ala Asp Met Gln Thr Arg Glu Val Leu Met Lys Leu Ser Gly	
	210 215 220	
5	act ata gag cgt gaa ggg ccc aac cgc cac ctc tat gac ttc act gga	720
	Thr Ile Glu Arg Glu Gly Pro Asn Arg His Leu Tyr Asp Phe Thr Gly	
	225 230 235 240	
10	aac ttg aac tta gat gga aaa gcc ttg ccc ttg ggc gct gac cag	768
	Asn Leu Asn Leu Asp Gly Lys Ala Leu Leu Pro Leu Gly Ala Asp Gln	
	245 250 255	
15	atc tta tta aga ggt aca cag ctt aga aat act cag tgg gtc ttt ggc	816
	Ile Leu Leu Arg Gly Thr Gln Leu Arg Asn Thr Gln Trp Val Phe Gly	
	260 265 270	
20	ata gtt gtt tat act gga cac gac acc aaa ctc atg cag aat tca acc	864
	Ile Val Val Tyr Thr Gly His Asp Thr Lys Leu Met Gln Asn Ser Thr	
	275 280 285	
25	aaa gca cct ctc aag aga tca aat gtt gag aag gtg act aac gtg cag	912
	Lys Ala Pro Leu Lys Arg Ser Asn Val Glu Lys Val Thr Asn Val Gln	
	290 295 300	
30	atc ctg gtg ttg ttt ggc atc ctc ttg gtc atg gcc ttg gtg agc tcg	960
	Ile Leu Val Leu Phe Gly Ile Leu Leu Val Met Ala Leu Val Ser Ser	
	305 310 315 320	
35	gcg ggg gcc ctg tac tgg aac agg tct cat ggt gaa aag aac tgg tac	1008
	Ala Gly Ala Leu Tyr Trp Asn Arg Ser His Gly Glu Lys Asn Trp Tyr	
	325 330 335	
40	atc aag aag atg gac acc acc tca gat aat ttt gga tac aac cta ctg	1056
	Ile Lys Lys Met Asp Thr Ser Asp Asn Phe Gly Tyr Asn Leu Leu	
	340 345 350	
45	acg ttc atc atc tta tac aac aat ctt att ccc atc agt ctg ttg gtg	1104
	Thr Phe Ile Ile Leu Tyr Asn Asn Leu Ile Pro Ile Ser Leu Leu Val	
	355 360 365	
50	act ctt gag gtt gtg aag tat act caa gcc ctt ttc ata aac tgg gac	1152
	Thr Leu Glu Val Val Lys Tyr Thr Gln Ala Leu Phe Ile Asn Trp Asp	
	370 375 380	
55	aca gat atg tat tat ata gga aat gac act cct gcc atg gcc agg aca	1200
	Thr Asp Met Tyr Tyr Ile Gly Asn Asp Thr Pro Ala Met Ala Arg Thr	
	385 390 395 400	
	tca aac ctt aat gaa gag ctt ggg cag gtg aaa tat ctg ttt tct gac	1248
	Ser Asn Leu Asn Glu Leu Gly Gln Val Lys Tyr Leu Phe Ser Asp	
	405 410 415	
	aag act gga acg ctt aca tgc aat atc atg aac ttt aag aag tgc agc	1296
	Lys Thr Gly Thr Leu Thr Cys Asn Ile Met Asn Phe Lys Lys Cys Ser	
	420 425 430	
	att gcc gga gta acc tat ggt cac ttc cca gaa ttg gca aga gag ccg	1344
	Ile Ala Gly Val Thr Tyr Gly His Phe Pro Glu Leu Ala Arg Glu Pro	
	435 440 445	

EP 1 225 182 A2

5	tct tca gat gac ttc tgt cgg atg cct cct ccc tgt agt gat tcc tgt Ser Ser Asp Asp Phe Cys Arg Met Pro Pro Pro Cys Ser Asp Ser Cys 450 455 460	1392
10	gac ttt gat gac ccc agg ctg ttg aag aac att gag gat cgc cat ccc Asp Phe Asp Asp Pro Arg Leu Leu Lys Asn Ile Glu Asp Arg His Pro 465 470 475 480	1440
15	aca gcc cct tgc att cag gag ttc ctc acc ctt ctg gcc gtg tgc cac Thr Ala Pro Cys Ile Gln Glu Phe Leu Thr Leu Leu Ala Val Cys His 485 490 495	1488
20	acg gtt gtt cct gag aag gat gga gat aac atc atc tac cag gcc tct Thr Val Val Pro Glu Lys Asp Gly Asp Asn Ile Ile Tyr Gln Ala Ser 500 505 510	1536
25	tcc cca gat gaa gct gct ttg gtg aaa gga gct aaa aag ctg ggc ttt Ser Pro Asp Glu Ala Ala Leu Val Lys Gly Ala Lys Lys Leu Gly Phe 515 520 525	1584
30	gtc ttc aca gcc aga aca cca ttc tca gtc atc atc gaa gcg atg gga Val Phe Thr Ala Arg Thr Pro Phe Ser Val Ile Ile Glu Ala Met Gly 530 535 540	1632
35	cag gaa caa aca ttc gga atc ctt aat gtc ctg gaa ttt tct agt gac Gln Glu Gln Thr Phe Gly Ile Leu Asn Val Leu Glu Phe Ser Ser Asp 545 550 555 560	1680
40	aga aaa aga atg tct gta att gtt cga act cct tca gga cga ctt cgg Arg Lys Arg Met Ser Val Ile Val Arg Thr Pro Ser Gly Arg Leu Arg 565 570 575	1728
45	ctt tac tgt aaa ggg gct gat aat gtg att ttt gag aga ctt tca aaa Leu Tyr Cys Lys Gly Ala Asp Asn Val Ile Phe Glu Arg Leu Ser Lys 580 585 590	1776
50	gac tca aaa tat atg gag gaa aca tta tgc cat ctg gaa tac ttt gcc Asp Ser Lys Tyr Met Glu Glu Thr Leu Cys His Leu Glu Tyr Phe Ala 595 600 605	1824
55	acg gaa ggc ttg cgg act ctc tgt gtg gct tat gct gat ctc tct gag Thr Glu Gly Leu Arg Thr Leu Cys Val Ala Tyr Ala Asp Leu Ser Glu 610 615 620	1872
60	aat gag tat gag gag tgg ctg aaa gtc tat cag gaa gcc agc acc ata Asn Glu Tyr Glu Glu Trp Leu Lys Val Tyr Gln Glu Ala Ser Thr Ile 625 630 635 640	1920
65	ttg aag gac aga gct caa cgg ttg gaa gag tgt tac gag atc att gag Leu Lys Asp Arg Ala Gln Arg Leu Glu Glu Cys Tyr Glu Ile Ile Glu 645 650 655	1968
70	aag aat ttg ctg cta ctt gga gcc aca gcc ata gaa gat cgc ctt caa Lys Asn Leu Leu Leu Gly Ala Thr Ala Ile Glu Asp Arg Leu Gln 660 665 670	2016
75	gca gga gtt cca gaa acc atc gca aca ctg ttg aag gca gaa att aaa Ala Gly Val Pro Glu Thr Ile Ala Thr Leu Leu Lys Ala Glu Ile Lys 675 680 685	2064
80	ata tgg gtg ttg aca gga gac aaa caa gaa act gcg att aat ata ggg	2112

EP 1 225 182 A2

	Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr Ala Ile Asn Ile Gly			
	690	695	700	
5	tat tcc tgc cga ttg gta tcg cag aat atg gcc ctt atc cta ttg aag		2160	
	Tyr Ser Cys Arg Leu Val Ser Gln Asn Met Ala Leu Ile Leu Lys			
	705	710	715	720
10	gag gac tct ttg gat gcc aca agg gca gcc att act cag cac tgc act		2208	
	Glu Asp Ser Leu Asp Ala Thr Arg Ala Ala Ile Thr Gln His Cys Thr			
	725	730	735	
	gac ctt ggg aat ttg ctg ggc aag gaa aat gac gtg gcc ctc atc atc		2256	
	Asp Leu Gly Asn Leu Leu Gly Lys Glu Asn Asp Val Ala Leu Ile Ile			
	740	745	750	
15	gat ggc cac acc ctg aag tac gcg ctc tcc ttc gaa gtc cgg agg agt		2304	
	Asp Gly His Thr Leu Lys Tyr Ala Leu Ser Phe Glu Val Arg Arg Ser			
	755	760	765	
20	ttc ctg gat ttg gca ctc tcg tgc aaa gcg gtc ata tgc tgc aga gtg		2352	
	Phe Leu Asp Leu Ala Leu Ser Cys Lys Ala Val Ile Cys Cys Arg Val			
	770	775	780	
25	tct cct ctg cag aag tct gag ata gtg gat gtg gtg aag aag cgg gtg		2400	
	Ser Pro Leu Gln Lys Ser Glu Ile Val Asp Val Val Lys Lys Arg Val			
	785	790	795	800
	aag gcc atc acc ctc gcc atc gga gac ggc gac aac gat gtc ggg atg		2448	
	Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Gly Met			
	805	810	815	
30	atc cag aca gcc cac gtg ggt gtg gga atc agt ggg aat gaa ggc atg		2496	
	Ile Gln Thr Ala His Val Gly Val Gly Ile Ser Gly Asn Glu Gly Met			
	820	825	830	
35	cag gcc acc aac aac tcg gat tac gcc atc gca cag ttt tcc tac tta		2544	
	Gln Ala Thr Asn Asn Ser Asp Tyr Ala Ile Ala Gln Phe Ser Tyr Leu			
	835	840	845	
	gag aag ctt ctg ttg gtt cat gga gcc tgg agc tac aac cgg gtg acc		2592	
	Glu Lys Leu Leu Leu Val His Gly Ala Trp Ser Tyr Asn Arg Val Thr			
	850	855	860	
40	aag tgc atc ttg tac tgc ttc tat aag aac gtg gtc ctg tat att att		2640	
	Lys Cys Ile Leu Tyr Cys Phe Tyr Lys Asn Val Val Leu Tyr Ile Ile			
	865	870	875	880
45	gag aaa att tag		2652	
	Glu Lys Ile *			

50	<210> 7		
	<211> 44		
	<212> PRT		
	<213> Artificial Sequence		
55	<220>		
	<223> consensus		
	<400> 7		

EP 1 225 182 A2

5 Pro Pro Thr Ala Thr Val Leu Arg Asp Gly Lys Glu Glu Glu Ile Pro  
1 5 10 15  
Ala Glu Glu Leu Val Pro Gly Asp Ile Val Glu Val Lys Pro Gly Asp  
20 25 30  
Arg Val Pro Ala Asp Gly Arg Val Val Glu Gly Glu  
35 40

10 <210> 8  
<211> 221  
<212> PRT  
<213> Artificial Sequence

15 <220>  
<223> consensus

20 <400> 8  
Thr Asn Leu Lys Ile Arg Gln Ala Leu Pro Glu Thr Gln Asn Leu Leu  
1 5 10 15  
Lys Glu Glu Asp Ser Phe Ser Pro Ser Asn Phe Asn Gly Glu Ile Glu  
20 25 30  
Cys Glu Lys Pro Asn Lys Asn Leu Tyr Ser Phe Gln Gly Asn Leu Thr  
35 40 45  
Leu Asn Asp Asp Glu Lys His Leu Glu Gly Ser Lys Gly Asp Gln Gly  
50 55 60  
Arg Arg Val Pro Leu Ser Pro Asp Asn Ile Leu Leu Arg Gly Cys Val  
65 70 75 80  
Leu Lys Asn Thr Glu Trp Val Tyr Gly Val Val Val Tyr Thr Gly His  
85 90 95  
Asp Thr Lys Val Met Met Asn Ser Ser Lys Pro Pro Arg Ser Lys Arg  
100 105 110  
Ser Arg Ile Glu Lys Glu Met Asn Met Met Ile Ile Ile Leu Phe Cys  
115 120 125  
Ile Leu Ile Val Leu Cys Leu Ile Ser Ala Ile Gly Ser Gly Ile Trp  
130 135 140  
Thr Arg Lys His Ser Lys Asn Arg Pro Tyr Tyr Ser Ser Pro Trp Tyr  
145 150 155 160  
Leu Lys Ser Asp Tyr Asn Pro Lys Asn Pro Ala Val Ser Gly Phe Val  
165 170 175  
Ser Phe Phe Thr Phe Ile Ile Leu Phe Ser Asn Leu Ile Pro Ile Ser  
180 185 190  
Leu Tyr Val Asn Ile Glu Ile Val Lys Ile Ile Gln Ser Tyr Phe Ile  
195 200 205  
Asn Trp Asp Leu Asp Met Tyr His Glu Glu Thr Asp Thr  
210 215 220

45 <210> 9  
<211> 136  
<212> PRT  
<213> Artificial Sequence

50 <220>  
<223> consensus

55 <400> 9  
Gln Asp Gln Ala Leu Arg Glu Lys Thr Gln Glu His Leu Glu Glu Tyr  
1 5 10 15  
Ala Arg Glu Gly Leu Arg Thr Leu Cys Ile Ala Tyr Arg Glu Leu Ser  
20 25 30  
Glu Glu Glu Tyr Glu Glu Trp Asn Gln Lys Tyr Glu Glu Ala Lys Thr

EP 1 225 182 A2

35 40 45  
Ser Ile Thr Glu Asp Arg Glu Glu Lys Leu Glu Glu Val Ala Glu Glu  
50 55 60  
Ile Glu Lys Asp Leu Val Leu Leu Gly Ala Thr Gly Ile Glu Asp Lys  
65 70 75 80  
Leu Gln Asp Gly Val Pro Glu Thr Ile Glu Thr Leu Arg Asn Ala Gly  
85 90 95  
Ile Lys Ile Trp Val Leu Thr Gly Asp Lys Met Glu Thr Ala Ile Asn  
100 105 110  
Ile Gly Tyr Ser Cys Lys Leu Ile Ser Arg Asn Met Lys Leu Ile Val  
115 120 125  
Ile Asn Glu Glu Ser Thr Asp Ser  
130 135

15  
<210> 10  
<211> 86  
<212> PRT  
<213> Artificial Sequence

20  
<220>  
<223> consensus

<400> 10  
Glu Val Gln Glu Leu Glu Thr Lys Ser Arg Asp Pro Gly Lys Ala Val  
25 1 5 10 15  
Leu Arg Asp Ser Asn Gly Lys Arg Leu Thr Glu Arg Ala Arg Leu Leu  
20 25 30  
Lys Asn Val Phe Arg Lys Arg Pro Asn Thr Leu Tyr Arg Thr Asp Ser  
35 40 45  
Leu Gln Gln Asn Leu Ser His Gly Tyr Ala Phe Ser Gln Glu Glu Asn  
30 50 55 60  
Gly Ala Val Pro Gln Ser Glu Ile Ile Arg Ala Tyr Asp Thr Thr Lys  
65 70 75 80  
Gln Arg Ser Arg Lys Lys  
85

35  
<210> 11  
<211> 7  
<212> PRT  
40 <213> Artificial Sequence

<220>  
<223> consensus

45  
<221> VARIANT  
<222> 6  
<223> Residue at position 6 also can be Leu, Ile, Val, or  
Met.

50  
<221> VARIANT  
<222> 7  
<223> Residue at position 7 also can be Thr or Ile.

<400> 11  
Asp Lys Thr Gly Thr Xaa Xaa  
55 1 5

## Claims

1. An isolated nucleic acid molecule selected from the group consisting of:
  - 5 a) a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6;
  - b) a nucleic acid molecule comprising a fragment of at least 120 nucleotides of the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6;
  - c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, or SEQ ID NO:5;
  - 10 d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5; wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, or SEQ ID NO:5; and
  - e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5; wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO: 1, 3, 4, 6, or a complement thereof, under stringent conditions.
2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:
  - 20 a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:6; and
  - b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, or SEQ ID NO:5.
- 25 3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences, or the nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.
4. A host cell which contains the nucleic acid molecule of claim 1.
- 30 5. The host cell of claim 5 which is a mammalian host cell.
6. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.
- 35 7. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a complement thereof.
  - 40 b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a complement thereof under stringent conditions; and
  - c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2 or SEQ ID NO:5.
- 45 8. The isolated polypeptide of claim 7 comprising the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:5, or the polypeptide of claim 7 further comprising heterologous amino acid sequences.
9. An antibody which selectively binds to a polypeptide of claim 7.
- 50 10. A method for producing a polypeptide selected from the group consisting of:
  - a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5;
  - b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, or SEQ ID NO:5; and
  - 55 c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or SEQ ID NO:5, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, or a complement thereof

under stringent conditions comprising culturing the host cell of claim 4 under conditions in which the nucleic acid molecule is expressed.

5       **11.** A method for detecting the presence of a polypeptide of claim 7 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 7; and  
b) determining whether the compound binds to the polypeptide in the sample.

10      **12.** The method of claim 11, wherein the compound which binds to the polypeptide is an antibody.

15      **13.** A kit comprising a compound which selectively binds to a polypeptide of claim 7 and instructions for use, or a kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

20      **14.** A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and  
b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

25      **15.** The method of claim 14, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

30      **16.** A method for identifying a compound which binds to a polypeptide of claim 7 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 7 with a test compound; and  
b) determining whether the polypeptide binds to the test compound.

35      **17.** The method of claim 16, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;  
b) detection of binding using a competition binding assay;  
c) detection of binding using an assay for 67073-mediated signal transduction.

40      **18.** A method for modulating the activity of a polypeptide of claim 7 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 7 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

45      **19.** A method for identifying a compound which modulates the activity of a polypeptide of claim 7, comprising:

a) contacting a polypeptide of claim 7 with a test compound; and  
b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

50

55

EP 1 225 182 A2

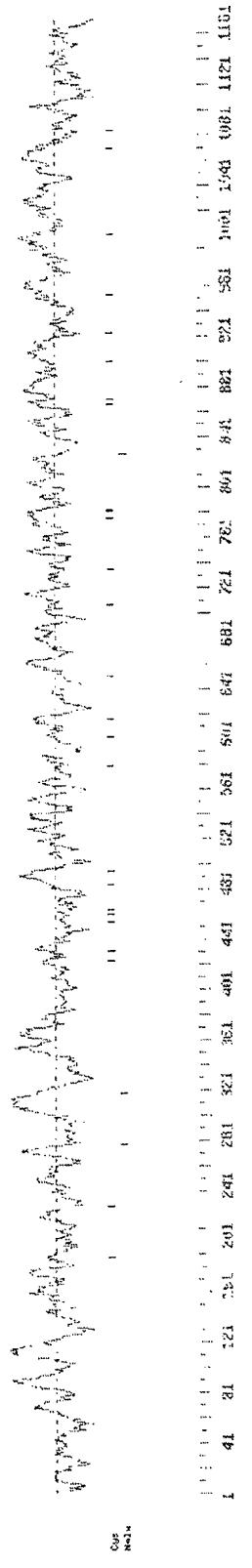


FIGURE 1